Addressing factors limiting the process efficiency of Pichia pastoris for the production of recombinant proteins

Kendrick, Emanuele

Award date:
2016

Awarding institution:
University of Bath

Addressing factors limiting the process efficiency of *Pichia pastoris* for the production of recombinant proteins

Emanuele Kendrick

A thesis submitted for the degree of Doctor of Philosophy

University of Bath
Department of Biology and Biochemistry
September 2016

Abstract

The methylotrophic yeast *Pichia pastoris* is a popular expression platform for the production of heterologous proteins, has been implemented in over 1500 laboratories and used to produce over 70 marketed proteins. However, despite high volumetric productivities achievable in fermentor culture, and the methanol inducible *AOX1* promoter, which can be manipulated to drive high levels of gene expression, *P. pastoris* suffers from low levels of cell specific productivity for secreted recombinant proteins. This has been attributed to bottlenecks in the secretory pathway, leading to product loss through the activation of ER stress responses, as well detrimental effects caused during cell growth on methanol. In addition a large disparity in protein production is observable within supposedly homogenous clonal populations raised from single cell lines, known as clonal variation, requiring extensive screening of transformants to find adequate producers.

The widespread use of zeocin, a known mutagen, as a selective agent was initially evaluated as a potential contributing factor to clonal variation. A study comprising data for the production of 4 different recombinant proteins from clonal samples raised in the presence of selection revealed that growth on zeocin does not exacerbate clonal variation and that selection itself appears to have a stabilising effect on the distribution of productivity phenotypes within certain populations. Variation in a range of native factors was also explored for a recombinant strain secreting *Penicillium funiculosum* glucose oxidase, including transgene transcription, transcription of the UPR marker *HAC1* and growth rate in culture. Whilst considerable variation was observed within samples for each of the factors tested, only a weak correlation between growth and productivity could be found, suggesting the involvement of other, currently unknown, factors that contribute to clonal variation.

The homologue of the yeast transcription factor – *OPI1* was deleted in *P. pastoris* as a strategy to increase productivity, as its removal in *S. cerevisiae* yields mutants with expanded ERs and improved secretion of recombinant proteins. However the loss of *OPI1* was shown to be detrimental to protein secretion, resulting in reduced productivities and an increased activation of the UPR. In a final study, 3 strains centered on the concept of uncoupling the activity of the Mut pathway transcription factor Mit1 from its transcriptional regulation were developed as a strategy to reduce glucose repression of the *AOX1* promoter. Each strain displayed alleviation of glucose repression to different degrees, with the least repressed strain - NGmit1; producing 20% more secreted recombinant protein than the native host during microexpression trials with a mixed feed of methanol and glucose.

Acknowledgements

First and foremost I would like to express my sincerest gratitude to my supervisor Professor David Leak for offering me the opportunity to become a part of his group and take on a PhD project with him. Throughout my entire course at the University of Bath David has been fair to me, on side, available to help troubleshoot my problems and has always valued my input in deciding the direction of this project. It has been a pleasure to be a part of what may be one David's final chapters in *Pichia* research (although it probably won't be because this is academia). My thanks also go to my second supervisor Dr Paul Whitley who has constantly made time, often at short notice, to meet with me, provided good advice, and always shown a genuine interest in my work. I would like to acknowledge all the efforts made by Dr Ian Hodgson as my supervisor at Fujifilm Diosynth as well as the guidance and intellectual input he's given throughout the course of my project. Special thanks also go to Dr Bryn Edwards-Jones and Gail Elgey for their support, input and who both, at different points, supervised my work at Fujifilm Diosynth and showed me how to work with 50 samples at any one time.

To all lab members of 1.28 and 1.33 as well as David's former group, past and present for my time in the lab - including in no particular order: Dr Jeremy Bartosiak-Jentys, Dr Rochelle Aw, Dr Kate Royle, Dr Ali Hussein, Dr Shyam Masakapalli, Dr Beata Lisowska, Dr Chris Ibenegbu, Lisa Buddrus, Micaela Chacon, Dr Matthew Styles, Dr Leann Bacon, Dr Charlie Hamley-Bennett, Dr Steve Bowden, Dr Chris Hills, Dr Alice Marriott, Mercedes Smith, Maria Vittoria Ortenzi, Ed Nesbitt, Daniel Baxter, Dr Anna Panek and Alex Holland. Many of you helped me get on my feet when I first started and maybe a few more times after that if you had to press me on it. I couldn't have asked for better colleagues. I would especially like to thank Natalie Vaughan for assisting me with my brief foray into confocal microscopy. I promise I will get you that McDonalds. To my parents: my father who I hold responsible for giving me a fascination in biology and my mother for her untiring support and determination to do anything in her power to try to help me through the most difficult times I've had to face these last few years. I know that I will always have her unending support in anything I choose to do and that everything in my life will be better for it. And to my partner Micaela who has always believed that I'm better than the person I am; thank you for being with me throughout and giving me hope every time I've felt useless or rubbish – I couldn't have finished this without you.

Contents

Figures list

Tables list

Abbreviations

µ	Micro
µg	Microgram
µl	Microlitre
Ω	Ohm
°C	Degrees celsius
aa	amino acids
AOX1	Alcohol oxidase 1
amp^r	Ampicillin resistance
BMG	Buffered minimal glycerol medium
BMGY	Buffered glycerol-complex medium
BMM	Buffered minimal methanol medium
BMMY	Buffered methanol-complex medium
bp	Base pair
CCR	Carbon catabolite repression
CHO	Chinese hamster ovary
CRISPR	Clustered regularly interspersed short palindromic repeats
CV	Coefficient of variation
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum associated degradation
g	Gram
GAP	Glyceraldehyde 3-phosphatase
GOx	Glucose oxidase
HIS4	Multifunctional protein involved in histidine biosynthesis in yeast
HSA	Human serum albumin
HuL T70A	Synthetic human lysozymecontaining the T70A amino acid change
Ire1p	Inositol requiring enzyme 1
h	Hour
KAR2	A karyogamy gene in yeast
kb	Kilobase
l	Litre
LB	Lysogeny broth medium
M	Molar
MD	Minimal dextrose medium
mg	Milligram
mol	Moles
mRNA	Messenger ribonucleic acid
Mxr1t	Functional, truncated form of Mxr1p consisting of its N-terminal 400aa sequence
nm	nanometres
OD	Optical density
P_{AOX1}	*AOX1* promoter
P_{GAP}	*GAP* promoter

PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
Pdi	Protein disulphide isomerase
pH	Measure of acidity or basicity in aqueous solution
rpm	Revolutions per minute
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
rDNA	Ribosomal DNA
SDS PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SOC	Super optimal broth with catabolite repression
TAE	Tris-acetate-EDTA buffer
U	Unit
UPR	Unfolded protein response
UPRE	Unfolded protein response element
UV	Ultra violet
V	Volt
v/v	Volume/ volume
w/v	Weight/ volume
YPD	Yeast extract, peptone, dextrose medium
zeo	Zeocin
zeoR	Zeocin resistance

1. Introduction

1.1 The industrial impact of recombinant protein production

Since human insulin was successfully expressed in *Escherichia coli* and subsequently manufactured as the first commercial recombinant product in 1982 (Ladisch and Kohlmann, 1992), recombinant protein production has formed the foundation of a global, multibillion-dollar enterprise. The advent of genetic engineering has allowed the manipulation of organisms into cell factories capable of producing high quantities of foreign proteins, through the insertion of heterologous genes into the genome of the host, and has revolutionised the manufacture of proteins in industry and academia. Until the latter part of the 20th century their molecular and structural complexity limited protein production to purification from cells of the native host, and the transition to the use of microorganisms as *in vivo* expression systems in large scale fermentations has now provided a more viable and cost effective alternative for the mass production of specific proteins (Hayworth, 2013). Moreover, through the cloning of sequence-modified target genes, expression systems have provided a platform for the genetic engineering of recombinant proteins diverging from their native homologues. This practice has not only proved essential in the discovery of novel, improved protein-based therapeutics, such as the generation of monoclonal antibody libraries, but also has been a key tool in researching the structure and function of important proteins (Chen, 2012). The global market for products made by recombinant expression technologies continues to show rapid growth, and is divided in the major part into two large niches: the manufacture of industrial enzymes and biopharmaceuticals.

1.1.1 The industrial enzymes market

As biological catalysts by nature, recombinantly produced enzymes across the tree of life have been successfully applied to a variety of chemical processes within vastly different industries to increase efficiency, product yields or refine final products. The sale of degradative enzymes such as lipases, carbohydrases and proteases dominate the market, with proteases alone being estimated to account

for ~60% of the total market share (Adrio and Demain, 2014). These are included in laundry detergents as well as being used in leather production, brewing and to improve the nutritional value of animal feed. Amylases make up the second most significant product by demand and are used in the degradation or modification of starch oligosaccharides within the paper, textiles, food and beverage industries (de Souza and de Oliveira Magalhães, 2010). Whilst its intrinsic links to such large, established global industries explain the continuing success of the industrial enzymes markets the advances and broadening scope within the field of biotechnology has also contributed to its steady growth (Adrio and Demain, 2014). The emergence of newer ventures, such as the production of 2nd generation biofuels and the biorefining of waste materials into value added chemicals has placed a greater demand for specialised enzymes capable of degrading more complex feedstocks at high temperatures to be able to operate within current industrial processes (Adrio and Demain, 2014). The global industrial enzymes market was valued at 4.62 billion USD as of 2015 with a projected annual growth of ~7.0% until 2024; predicted to be driven by the expansion of the leather, bioethanol and nutraceutical industries (Grand View Research, 2016).

1.1.2 The biopharmaceutical market

Biopharmaceuticals, often shortened to biologics, encompass all drugs used to treat human disease of which either all or a component is made up of a biological macromolecule. Biologics represent by far the largest market for recombinant products, being valued at 162 billion USD in 2014 and expected to rise to 278 billion USD by the year 2020 (Persistence Market Research, 2015). This reflects the impressive pipeline of novel biologics currently in development by pharmaceutical companies responsible for a record of 20 new, FDA-approved drugs being released in 2015 alone (U.S. Food and Drug Administration, 2015). The rapid adoption of biologics is evident by the growth in their share of the global pharmaceutical market from just 10% in 2000 to 29% in 2014, and accounting for over 50% of the pharmaceutical revenue streams of top companies: Sanofi-Aventis, Roche and AbbVie in 2012 (Otto, Santagostino and Schrader, 2014). Within the biologics market itself therapeutic monoclonal antibodies make up the dominant class of product and represented almost half of total, global sales, netting close to 75 billion USD in 2013 (Ecker, Jones and Levine, 2015; Kinch, 2015). A greater understanding of the molecular interactions governing antibody binding and antigen recognition has led to the engineering of monoclonal antibodies with high specificity and drug

delivery properties to a wide range of disease targets, or otherwise, that are often greatly improved over their small molecule drug counterparts (Ecker, Jones and Levine, 2015). As such monoclonal antibody products have been implemented in a variety of medical applications from pregnancy tests to diagnostics to the treatment of cancers and autoimmune diseases.

For their many advantages, a major drawback to recombinant proteins lies in the cost of their production, especially for biologics which are more heavily regulated due to the inherent nature of their end-use (Bren, 2006). These costs are passed on from companies to the consumer, raising serious implications for the access of affordable services and products, particularly in the case of healthcare in a time where an increasing number of drugs used to treat disease are based on recombinant proteins. For example in the United Kingdom the therapeutic monoclonal antibody Kadcyla®, made by Roche, was rejected by the National healthcare system as a routine prescription for breast cancer, citing its list price of 90,000 GBP per patient as the main reason (National Institute for Health and Care Excellence, 2015). Hence investments have been made in advancing upstream and downstream technologies to maximise the efficiency of production in an effort to drive down final product costs while maintaining the profit margins of the companies involved.

1.2 Microbial expression systems

From a similar bioprocessing perspective it would also be pertinent to increase the efficiency and productivity of expression systems in order to reduce manufacturing costs and increase the yield of high value recombinant proteins. Furthermore proteins developed for clinical use often require more complex cellular machinery typical to eukaryotes to be correctly folded and processed, such as the formation and breakage of disulphide bonds between cysteine residues, as well as a pattern of post-translational modifications that will render the protein stable and functional without eliciting an immune response when administered to humans (Schmidt, 2004). The choice of microorganisms as hosts for recombinant protein production benefits from a number of advantages such as robust growth on inexpensive media and a high turnover of product (Meyer and Schmidhalter, 2012). The ability to conduct protein production from microbial hosts in large scale fermentations within bioreactors also minimises space requirements and the equipment footprint for the process over more complex hosts such as plants or transgenic animals. Extensive research into genetic tools, strain

development, media composition and fermentation strategies has resulted in the availability of a wide range of well characterised expression platforms from a variety of microbial hosts that have been optimised for the production of different target proteins. Perfect, universal production of all commercial recombinant proteins meeting the correct specification of the final product is impossible in any one expression system, and has led to the propagation of multiple systems that occupy different niches of product requirements – as exemplified by the range of hosts used by the biopharmaceutical industry within this decade (Figure 1).

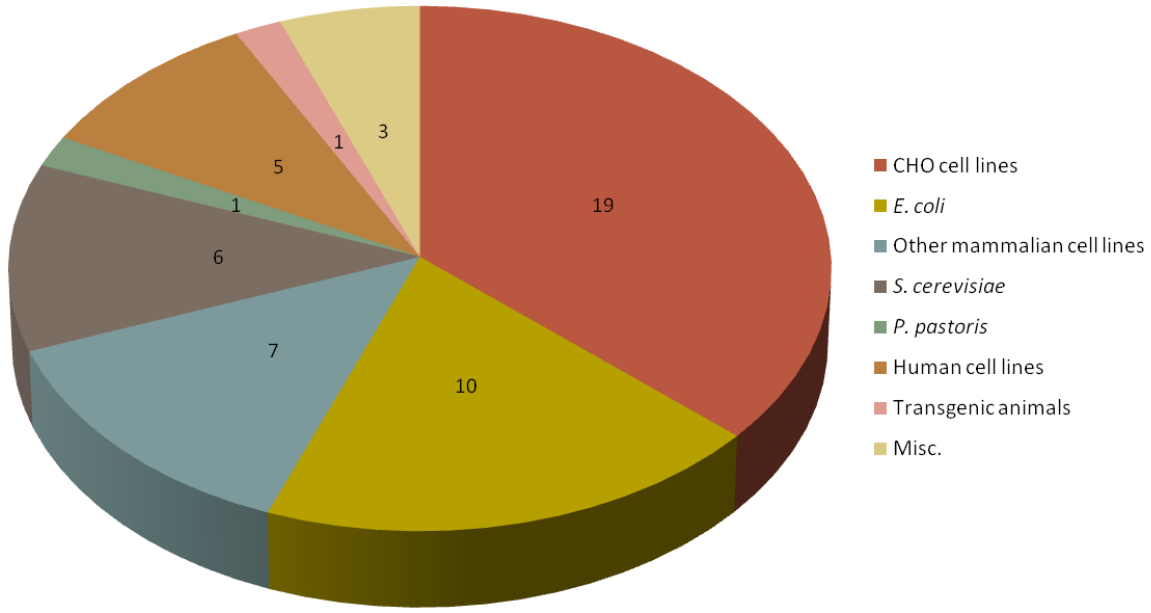


Figure 1: Data adapted from Walsh (2014) to display the distribution of expression hosts used in the production of all approved biopharmaceutical recombinant proteins between 2010 and 2014. Mammalian cell and *E. coli* expression systems continue to produce the clear majority of biologics although other notable hosts include the yeasts *S. cerevisiae* and *P. pastoris* that have been used in the successful production of new biological entities that have passed clinical trials.

1.2.1 Bacterial expression systems

Prokaryotic expression hosts currently remain the most popular systems for heterologous protein production as their ease of genetic manipulation, simple growth requirements and high levels of

protein expression have made them widely accessible (Chen, 2012). A workhorse for heterologous protein production in research and many industrial sectors is the model organism *E. coli*, which has fast established itself as arguably the most popular expression system (Rosano and Ceccarelli, 2014). *E. coli* is capable of growing to high cell densities of up to 200g dry cell weight per ml of culture (Lee, 1996) and, under optimal conditions, can grow at doubling times of up to 20 minutes in glucose salts medium (Sezonov, Joseleau-Petit and D'Ari, 2007), often meaning that batch cultures can grow to stationary phase within less than a day after inoculation. In addition to its superior growth characteristics simple and efficient cloning techniques have been well established for the *E. coli* system (Pope and Kent, 1996; Froger and Hall, 2007) and a wealth of advanced molecular tools and engineered strains are available to optimise the expression of specific proteins (Chen, 2012).

An alternative host commonly used in the production of recombinant proteins is the Gram positive bacterium *Bacillus subtilis* which, unlike *E. coli*, lacks any cell surface endotoxic lipopolysaccharides (LPS), granting it full GRAS status (Terpe, 2006). Furthermore *B. subtilis* contains a high-capacity secretory pathway that can be manipulated to export recombinant proteins into the culture medium, aiding in product recovery (Simonen and Palva, 1993). The higher guarantee that purified, secreted recombinant proteins from *B. subtilis* will be endotoxin-free has made it the more favoured bacterial expression system within the food industry (Westers, Westers, and Quax, 2004).

1.2.2 Mammalian cell lines

Bacterial expression systems are limited in the size and complexity of the recombinant proteins that can be expressed (Braun, 2002; Khow and Suntrarachun, 2012), as they do not possess the cellular machinery for post-translational modifications that are present in higher eukaryotes. Mammalian cells are used extensively in the biopharmaceutical industry as their form of post-translational modification is highly similar to that of human cells, minimising the risk of an immune response or loss of functionality of complex therapeutic proteins (Vink et al, 2013). Currently the most used and well studied mammalian expression systems are derived from immortalized Chinese hamster ovary (CHO) epithelial cell lines. The first approved mammalian cell-expressed therapeutic protein, tissue plasminogen activator, was produced in CHO cells in 1987 (Gaffney and Curtis, 1987) and since then favourable growth characteristics and successive strain and process improvements have established

CHO cells as the primary expression system within the biopharmaceutical industry (Walsh, 2014). Other notable expression systems include the murine NS50 and the human embryonic kidney (HEK) cell lines, which have seen an increase in usage in the production of biologics (Gaughan, 2016).

Nevertheless the costly media components required for growth, more complicated handling over bacterial systems and the higher risk of viral contamination arguably constrain the usage of mammalian expression systems to the production of very high value products within large industries, which fits in the case of biologics (Boehm, 2007; Yin et al, 2007). The time required to transfect and grow a stable recombinant cell line is also far longer in comparison to bacterial systems, further increasing the cost of production (Hacker, 2009).

1.2.3 Yeast expression systems

Yeast cells offer a number of advantages over both bacterial and mammalian expression hosts: prolific growth and expression can be achieved with simpler growth requirements and, as eukaryotes; they are capable of post-translational modifications (Nasser et al, 2003). *Saccharomyces cerevisiae* and, more recently, *Pichia pastoris* are the most commonly used yeast expression systems, with an extensive knowledge base and genetic toolbox developed from years of studying *S. cerevisiae* as a model organism (Nasser et al, 2003). The major disadvantage of *S. cerevisiae* as a yeast system is that the glycosylation profile it modifies proteins with differs significantly from mammalian cells, limiting their use for the production of human therapeutic proteins. N-linked glycosylation in human cells entails the removal of mannose from attached glycans and the addition of N-Acetylglucosamine (GlcNAc), galactose, fucose and sialic acid subunits, whereas *S. cerevisiae* is limited to the addition of further mannose and mannosylphosphate sugars (De Pourcq, De Schutter and Callewaert, 2010). The extensive addition of mannose residues, known as hyper mannosylation, results in the attachment of numerous large, complicated, glycan structures to secreted proteins and has been observed in many yeast systems (De Pourcq, De Schutter and Callewaert, 2010). Hyper-mannosylated glycoforms of recombinant proteins often exhibit reduced bioactivities and also hamper the scope for yeast production of biologics as they are rapidly cleared in the human bloodstream and can be immunogenic (Yip et al, 1994; Celik and Calık, 2012).

1.3 *Pichia pastoris* (*Komagataella phaffi*); a history

Pichia pastoris represents one of a number of genera of methylotrophic yeasts; the only eukaryotic organisms capable of utilising methanol as a sole carbon source (Yurimoto, Oku and Sakai, 2011). Owing to the high levels of biomass that could be achieved through growing on methanol, the organism was initially adopted for the production of single cell protein (SCP) during the 1970's by Phillips Petroleum. The low cost of methane, a major component for the industrial manufacture of methanol, made *P. pastoris* an attractive candidate for SCP production and so the first *P. pastoris*-defined media and fermentation protocols were designed (Cregg et al, 2000). However the rise in methane prices during the 1973 oil crisis led to the project becoming economically impractical until the development of yeast molecular techniques in the 1980's gave rise to the possibility for the use of *P. pastoris* as an expression system (Cregg et al, 2000; Macauley-Patrick et al, 2005). With the intention of creating high value SCP containing recombinant growth factors Philips Petroleum funded research at Salk Institute BioTechnology/Industrial Associates to create the first molecular tools for the genetic manipulation of *P. pastoris*, and to identify suitable promoters for the expression of heterologous proteins (Cregg et al, 1985; Biogrammatics, 2013). The combination of strong native promoters to drive expression with the scale-up production processes already developed during its SCP years finally realised the potential of *P. pastoris* as an effective expression system during early trials in the 1990's. A notable example includes the large-scale production of hydroxynitrile lysase, derived originally from tropical rubber trees, from which final titres exceeding 20g/l were achieved (Hasslacher et al, 1997). The first commercially available *Pichia* expression toolkits were released by Research Corporation Technologies (RCT), who have held the patent to the system since 1993 and have granted licenses to other companies, including Life Technologies, now part of Thermo Fisher Scientific, to increase its accessibility. Publications relating to the use of *P. pastoris* as an expression system rose exponentially throughout the 1990's following the system's commercialization to the point where over 200 heterologous proteins had been reported to have been expressed successfully by the year 2000 (Cregg, Cereghino and Higgins, 2000). The publication of the first full *P. pastoris* genome sequence in 2009 (De Schutter, 2009), the formation of new strains, tools and the transferability of many previously recognized yeast protocols have helped to establish *Pichia* as a mainstream, well-defined yeast expression system (Ahmad et al, 2014). Despite this *S. cerevisiae* clearly remains the primary yeast expression system within the biopharmaceutical industry (Figure 1)

and, as of 2014, only two currently approved commercialised biologics have been produced in *P. pastoris*: Jetrea – a truncated form of human plasmin by Thrombogenics (Belgium) and Kalbitor – a plasma kallikrein inhibitor produced by Dyax (USA). However the *Pichia* expression system has gained significant traction in other sectors and has been used to make over 70 commercial products (Research Corporation Technologies, 2013) and has even succeeded *S. cerevisiae* as the most cited yeast expression system for certain applications, such as the production of lipases (Valero, 2012).

Despite the phenotypic similarities originally observed across the methylotrophic yeasts in the *Pichia* genera that led to their initial grouping, sequencing studies of the 18S and 26S ribosomal RNAs revealed that the genus was not monophyletic; instead containing a number of distantly related species (Yamada, Maeda, and Mikata, 1994). The results substantiated the reclassification of a number of species including *P. pastoris*, which was assigned to the newly described genus *Komagataella* (Yamada et al, 1995). The result of a multigene sequence analysis between members of the *Komagataella* genus in 2009 revealed that *K. pastoris* was in fact synonymous with the species *K. phaffi*, described in 2005 (Kurtzman, 2005), and should therefore be reclassified as such (Kurtzman, 2009). Whilst the phylogenetic classification is formally recognised, it has not completely replaced *P. pastoris* as the associated name for the organism within the community with many academic articles published since the finding and throughout the period of this project continuing to reference *P. pastoris*. Even though seven years have passed since its reclassification the reluctance to fully replace all reference to *P. pastoris* with *K. phaffi* is most likely due to the fact that the fundamental research establishing the *Pichia* system and its commercialisation was all conducted under the name *P. pastoris*, proposing that a certain degree of brand recognition exists in the name. For the sake of clarity, the organism formally known as *K. phaffi* will also be referred to as *P. pastoris* within this project.

1.4 Key features of the *Pichia* system

1.4.1 Growth on methanol

The key feature that first garnered interest in *P. pastoris* as a candidate for use in biotechnology processes was its ability to reach high cell densities in minimal media containing methanol as a sole carbon source for respiration. Methanol is sequentially converted into glyceraldehyde 3-phosphate and assimilated into the tricarboxylic acid cycle (TCA) through a tightly regulated, unique pathway conserved among methylotrophic yeasts, named the methanol utilisation (Mut) pathway (Hartner and Glieder, 2006) (Figure 2).

The initial steps of the Mut pathway produce the highly oxidative species hydrogen peroxide and formaldehyde. To prevent oxidative damage to the cell interior these steps are housed in specialised peroxisomes, containing enzymes catalyzing the first three reactions of methanol assimilation, which proliferate from the endoplasmic reticulum (ER) upon growth on methanol and occupy large cytoplasmic spaces in the cell. Of the peroxisomal Mut pathway enzymes alcohol oxidase (AOX) catalyses the oxidation of methanol into formaldehyde and hydrogen peroxide, and is essential for methanol metabolism in *P. pastoris*. In what is possibly an evolutionary adaptation to compensate for the suboptimal kinetics of AOX caused by its low affinity for oxygen, *P. pastoris* produces large quantities of its primary alcohol oxidase (AOX1) during growth on methanol. Production of AOX1 is so high that it accumulates to form up to 30% of the total, soluble, intracellular protein content in methanol-grown cells (Couderc and Baratti, 1980) (Hartner and Glieder, 2006).

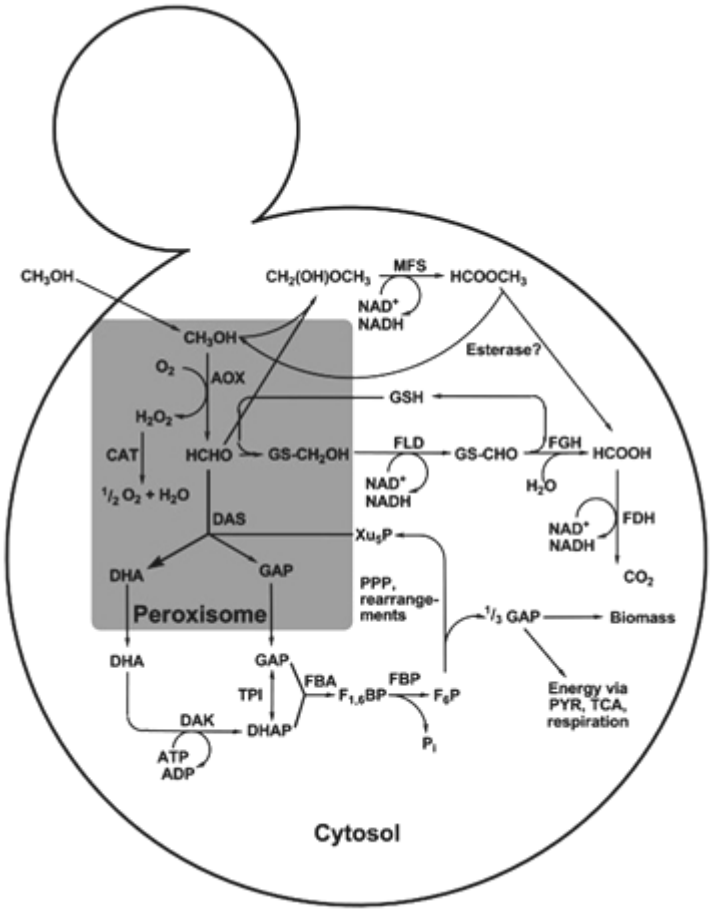


Figure 2: Schematic displaying the methanol utilisation pathway in *P. pastoris* from Hartner and Glieder (2006). Abbreviated enzymes/pathways/molecules are AOX: alcohol oxidase, CAT: catalase, FLD: formaldehyde dehydrogenase, FGH: S-formylglutathione hydrolase, FDH: formate dehydrogenase, DAS: dihydroxyacetone synthase, TPI: triosephosphate isomerase, DAK: dihydroxy-acetone kinase, FBA: fructose 1,6-bisphosphate aldolase, FBP: fructose 1,6-bisphosphatase, MFS: methylformate synthase; DHA: dihydroxyacetone, GAP :glyceraldehyde 3-phosphate, DHAP: dihydroxyacetone phosphate, $F_{1,6}BP$: fructose 1,6-bisphosphate, F_6P: fructose 6-phosphate, Pi: phosphate, Xu_5P: xylulose 5-phosphate, GSH: glutathione, PYR: pyruvate; PPP: pentose phosphate pathway, TCA: tricarboxylic acid cycle

The deletion of the *AOX1* gene results in mutants with reduced growth on methanol (MutS) while the double deletion of *AOX1* and the secondary, lesser expressed *P. pastoris* alcohol oxidase – *AOX2* (Koutz et al, 1989) causes the complete impairment of methanol metabolism (Mut-) (Chiruvolu, Cregg, and Meagher, 1997). Whilst the wild type phenotype (Mut+) for the Mut pathway is generally favoured as standard for expression in *P. pastoris*, the MutS phenotype holds bioprocess value as it

has been used in conjunction with modified fermentation strategies to increase production of certain recombinant proteins (Pla et al, 2006) (Krainer et al, 2012).

1.4.2 Inducible and constitutive promoters

Growing *P. pastoris* on methanol also serves a dual purpose as the Mut pathway plays host to a number of methanol-inducible promoters that regulate the expression of genes within it, and are therefore suitable for controlling transgene expression. Of these the *AOX1* promoter (P_{AOX1}), which is responsible for directing high levels of transcription of *AOX1* to meet its high production demand, is the most commonly used due to its characteristically high strength. P_{AOX1} is also strictly regulated by a number of *trans*-acting mechanisms in response to growth in different sources; expression is fully activated only with the presence of methanol in the culture medium and tight repression occurs during cell growth on glucose or glycerol. Its switchable nature has led to P_{AOX1} becoming a classic genetic tool for tuning transgene expression in *P. pastoris* and has allowed the development of fermentation strategies with discrete stages that prioritise cell growth, followed by protein production, proving particularly useful for the production of recombinant proteins with cytotoxic effects (Jahic et al, 2006). The strength of P_{AOX1} can be detrimental to the production of certain heterologous proteins if other downstream factors such as translation, folding or post-translational processing become limiting factors. Therefore weaker promoters of other methanol inducible genes comprising the Mut pathway have also been tested successfully, including the *AOX2*, *FLD1* and *DAS* and *PEX8* promoters. Methanol induction in itself is not always viable for certain processes due to the inherent safety concerns associated with working with large volumes of methanol. This has spurred the search for alternative inducible promoters and, in 2013, the glucose-regulated promoters P_{G1} and P_{G6} were identified through microarray analysis (Prielhofer et al, 2013). Both promoters are strongly repressed in cultures containing excess glucose, and are subsequently activated to induce high levels of transcription upon glucose depletion in culture (Prielhofer et al, 2013).

A number of promoters for constitutive transgene expression are also available in the *P. pastoris* genetic toolkit of which the P_{GAP} promoter, controlling the expression of the glyceraldehydes 3-phosphate dehydrogenase gene natively in *P. pastoris*, has emerged as the most popular alternative to P_{AOX1}. Advantages of P_{GAP} include strong expression when cells are grown in glucose or glycerol-

based media, resulting in high protein production which can compare favourably to P_{AOX1} under certain situations. As in the case of methanol induction, weaker constitutive promoters have also been validated and include the promoters for translation elongation factor 1 alpha (P_{TEF}), enolase (P_{ENO1}) and a GTPase (P_{YTP1}).

1.4.3 Common expression vector design

A regular challenge that presents itself in a number of expression systems is in the replication and maintenance of vectors containing the target transgene throughout the entire process of culture growth and expression. The recombinogenic nature of *P. pastoris* has allowed the design of vectors that are maintained stably in cells by integrating into specific locations in the host genome (Mokdad-Gargouri et al, 2012). Expression vectors containing regions sharing homology to the target locus in the genome are restricted to form a linear construct with homologous ends. Upon successful transformation into cells homologous recombination (HR) directs the insertion of the vector into the target locus through a single or double crossover event. Common sites of integration include the regions containing the native *P. pastoris* promoter included in the vector to direct transgene expression, such as the *AOX1* or *GAP* loci (Sears et al, 1998). Since relatively large quantities of vector are recommended for the efficient transformation of integrating vectors, with approximately 5-10µg total DNA content used as standard for a single transformation of *P. pastoris* by electroporation, bacterial origins of replication and selectable markers are included in *Pichia* system vectors to enable their storage and biosynthesis in *E. coli* hosts (Higgins and Cregg, 1998). A host of products employing this benchmark design are available commercially and include the pPICZα range of plasmids; a popular set of vectors for *AOX1*-based expression marketed by Thermo Fisher Scientific and used as the basis for a large number of the expression vectors used within this study (Figure 3).

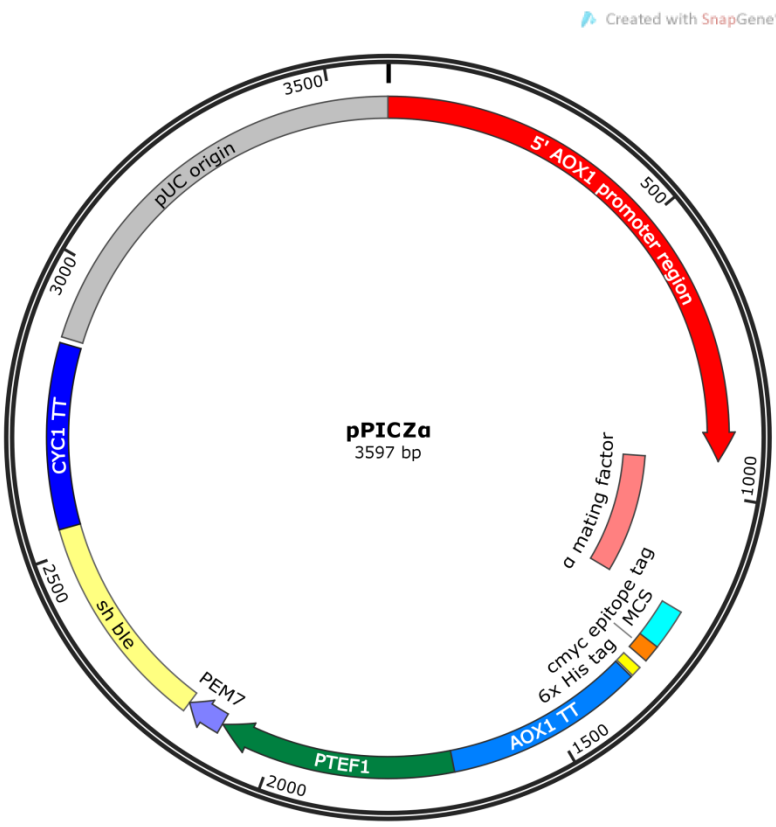


Figure 3: Vector map of pPICZα. Notable features include the *AOX1* promoter region for methanol inducible expression and integration into the *AOX1* locus, the *S. cerevisiae* α-mf for heterologous protein secretion, polyhistidine tag for purification and the sh ble selectable marker conferring resistance to zeocin.

Whilst the genomic integration of vectors into the genome persists as the most common method for maintaining stable expression, increasing interest in the value of high-throughput cloning techniques and transient expression for synthetic biology projects has spurred the further development of episomal vectors for use in *P. pastoris* (Camattari et al, 2016). An episomal plasmid containing a 452bp autonomous replicating sequence originating from *Kluyveromyces lactis* (panARS) was successfully assembled *in vivo*, cloned and retained in *P. pastoris* and even outperformed equivalent integrated vectors in the expression of blue fluorescent protein (Camattari et al, 2016). Whilst the use of ARS's in *P. pastoris* has been historically validated (Cregg, 1986) the field appears to be experiencing a resurgence, with exploratory studies of the *P. pastoris* genome in 2014 identifying a number of native ARS sequences from which to develop new episomal vectors (Liachko et al, 2014). Most recently a novel plasmid utilising native centromeric DNA to autonomously replicate in *P. pastoris* has also been described (Nakamura et al, 2016).

1.4.4 Post translational modifications

Whilst other standard yeast expression systems such as *S. cerevisiae* have been reported to hyper-mannosylate heterologous proteins during post-translational modification, a key feature that distinguishes *P. pastoris* from *S. cerevisiae* is that it is less prone to hyper-mannosylation, and so is more suited to the expression of secreted proteins (De Pourcq, De Schutter and Callewaert, 2010; Jahic, 2013). Furthermore the arrival of a glycoengineered, "humanized" strain of *P. pastoris* in 2006, capable of performing human type N-linked glycosylation, has potentially overcome the issues of the immunogenicity of therapeutic proteins expressed in yeasts (Hamilton, 2006). The scale-up production of a fully functional monoclonal antibody with a high uniformity of N-linked glycans, at a productivity of over 1g/l by a glycoengineered *P. pastoris* strain in 2009 (Potgeiter, 2009) has demonstrated that *P. pastoris* could also rival mammalian expression systems for the production of therapeutic proteins. However, in spite of the commercial availability of glycoengineered strains, the most notable being the *Pichia* Glycoswitch® system from RCT, the usage of *P. pastoris* within the biopharmaceutical industry has not increased in response. A potential reason is that the principles originally applied in the glycoengineering of *P. pastoris* were replicated in *S. cerevisiae* more recently to develop strains also capable of performing humanized N-linked glycosylation (Arico, Bonnet, and Javaud, 2013). The availability of two different yeast expression systems capable of uniform, human-like glycosylation and the longer history of *S. cerevisiae* in the biopharmaceutical industry could provide an explanation for its low adoption of glycoengineered *P. pastoris*.

1.4.5 Recombinant protein secretion

Heterologous proteins can be targeted for secretion in the *Pichia* system through addition of a number of yeast signal sequences, and benefit from a low background of endogenous proteins secreted by *P. pastoris* into the media (Cregg et al, 2000; Lin-Cereghino et al, 2002). Therefore, if cells are grown in minimal media containing no added protein, the secreted recombinant protein often constitutes the vast majority of the total protein in the medium after expression, reducing the cost and resources required in downstream processing. The most common signal sequence used to target

proteins to the secretory pathway in *P. pastoris* encodes the alpha mating factor (α-mf) from *S. cerevisiae*, and is typically cloned to be expressed at the N-terminus of the target protein (Kurjan, and Herskowitz, 1982). Other secretion signals have also been used in the past and include the signals for the *S. cerevisiae* invertase (*SUC2*) and the *P. pastoris* acid phosphatase (*PHO1*) (Daly and Hearn, 2005).

1.5 Advanced tools and techniques

1.5.1 Genome sequence

The first publication of a *P. pastoris* genome in 2009, specifically from the strain GS115, has provided information that greatly enabled strain improvement strategies and contributed to the associated knowledge base (De Schutter et al, 2009). Accessibility to the genome for users was improved significantly through its integration with a web-based Genome browser made publicly available; with a Wiki based platform to allow users to add new annotations for currently identified genes (Mattanovich et al, 2009). In 2011 a high quality genome sequence for the parental *P. pastoris* strain – CBS7435 was published and included, for the first time, sequence for the *P. pastoris* mitochondrial genome (Kuberl et al, 2011). The most ambitious sequencing project to date assembled genomes from seven different strains and developed an updated reference genome sequence, correcting over 500 positions from previous sequences and including sequences of putative chromosomal centromeres as well as two cytoplasmic plasmids found in certain strains (Sturmberger et al, 2016). The successive revisions and additions to the *P. pastoris* genome sequence has provided a high quality resource for the user community from which to mine for elements or genes that could exploited as new molecular biology tools or identify new potential targets for strain engineering projects.

1.5.2 Promoter libraries

The industrial and commercial relevance of the *AOX1* promoter for methanol inducible protein production has led to studies attempting to elucidate its transcriptional regulation (Vogl and Glieder 2013). Truncation studies as well as the investigation of predicted transcription factor binding sites resulted in the mapping of multiple *cis*-acting regulatory sequences (Inan, 2000; Hartner et al, 2008; Xuan et al, 2009). The sequential deletion or repetition of these elements by Hartner et al (2008) resulted in the development of a P_{AOX1} variant library ranging from ~6% to ~160% of the original promoter strength of P_{AOX1}, based on the expression of recombinant GFP. Considering that the strength of the native P_{AOX1} can somewhat tuned by controlling the concentration of methanol in culture (Jahic et al, 2006), the availability of a P_{AOX1} promoter library provides the user with more control over transgene expression levels to optimise the productivity of specific proteins, and has increased the versatility of methanol induction in *P. pastoris*.

Constitutive expression has been similarly innovated and, in 2011, a *GAP* promoter library was generated by error prone PCR and quantified for activity by measuring the associated expression of a yeast-enhanced GFP (Qin et al, 2011). A wide range of promoter activities were encompassed in the library with the strongest variant, named P_{G1}, increasing expression by 19.6 fold over the native P_{GAP}.

1.5.3 Post transformational vector amplification (PTVA)

As discussed previously, the recombinogenic nature of *P. pastoris* permits the design of vectors that target genes to integrate into a specific locus within the host genome by homologous recombination. Through this process multicopy clones can spontaneously arise through the integration of more than one copy of the vector into the genome, and in many cases has been shown to increase specific productivity through an increase in gene dosage (Clare et al, 1991). The importance of gene dosage in increasing productivity was discovered for a number of different proteins, leading to the development of techniques to enrich transformant populations for high copy number variants, or to stimulate copy number increase *in vivo* (Cos et al, 2005). In 2008 Sunga et al described a protocol to select for multicopy integrants through the re-plating of transformant colonies containing vectors bearing the zeocin resistance marker on agar media containing successively higher concentrations of

zeocin. The study reported that the use of PTVA was able to increase the probability of obtaining a "jackpot" colony containing 10 or more vector copies from ~1% to 5-6%, and has been generally adopted as a process for selecting multicopy integrants of *P. pastoris* (Sunga et al, 2008). Since its conception PTVA has also been validated for selection with G418 and hygromycin, the latter being significantly less expensive than zeocin (Cole, 2009; Yang et al, 2014). Recently the protocol was adapted for the selection of strains in liquid media, significantly reducing antibiotic usage and clone growth times for the zeocin and G418 PTVA protocols (Aw and Polizzi, 2016).

1.5.4 Improved platform strains for genetic engineering and productivity

In conjunction to the those previously discussed, considerable research into strain development has given rise to a range of engineered platforms, many of which have been made available commercially, that improve on the usability or efficacy of the *Pichia* system. A compilation of some of the more notable strain modifications made in *P. pastoris*, and their outputs are summarised in Table 1.

The initial development of genetic tools to engineer *P. pastoris* resulted in the creation of auxotrophic mutant strains incapable of synthesising different organic compounds, allowing for the selection of positive transformants with vectors containing the required gene to rescue the biosynthetic pathway (Cregg et al, 1985; Lin Cereghino et al., 2001). Many of these are still used as popular expression platforms, including *P. pastoris* GS115, currently marketed by Thermo Fisher Scientific and Ade- strains, which are included within the *Pichia*Pink™ commercial expression kit. Also included within *Pichia*Pink™ are strains lacking the major proteases Pep4 and proteinase B, which have been demonstrated to improve the production of certain recombinant proteins that are particularly sensitive to protease-mediated degradation (Gleeson et al, 1998; Ahmad et al, 2014). The further improvement of production of different recombinant proteins has been achieved in engineered strains achieving higher rates of secretion, with common strategies for strain improvement centring on overexpressing factors assisting with the folding and processing of proteins within the secretory pathway (5.1).

.

Table 1: Compilation of modifications conducted to generate key platform *P. pastoris* strains

Modification	Description	Output	Reference
Methanol utilisation mutants			
aox1	Disruption of the gene encoding alcohol oxidase 1 (AOX1), generating the MutS phenotype.	Slow growth on methanol. Increased production of A33 scFv fragment over the Mut+ phenotype	Cregg and Madden (1987) Pla et al (2006)
aox1, *aox2*	Disruption of both alcohol oxidase genes, generating the Mut- phenotype.	No growth on methanol.	Chiruvolu, Cregg, and Meagher (1997)
Auxotrophic mutants			
his4	Disruption of *HIS4*, encoding a multifunctional enzyme essential for histidine biosynthesis.	Histidine auxotrophic strain GS115 (Thermo Fisher Scientific). Selectable by *HIS4* complementation.	Cregg et al (1985)
ade2	Disruption of *ADE2*, encoding a Phosphoribosylaminoimidazole carboxylase - essential for adenine biosynthesis.	Adenine auxotrophic strain JC220. Selectable by *ADE2* complementation.	Lin Cereghino et al., 2001
arg4	Disruption of *ARG4*, encoding Argininosuccinate lyase – essential for arginine biosynthesis.	Arginine auxotrophic strain GS190. Selectable by *ARG4* complementation.	Lin Cereghino et al., 2001
ura3	Disruption of URA3, encoding Orotidine 5'-phosphate decarboxylase – essential for uracil biosynthesis.	Uracil auxotrophic mutant JC254. Selectable by *URA3* complementation.	Lin Cereghino et al., 2001
Glycoengineered strains			
och1	Disruption of the α-1,6-mannosyltransferase gene in *P. pastoris* to reduce hyper-mannosylation.	Incorporated into the *Pichia* Glycoswitch® expression platform. Contains a range of strains and plasmids expressing different combinations of the stated glycosyltransferases to perform more uniform, as well as mammalian type N-glycosylation of recombinant proteins.	Vervecken et al (2007); Laukens et al (2015)
H. sapiens GNT-I	Expression of Human GlcNAc transferase I.		
R. norvegicus GNT-II	Expression of rat mannosidase II.		
H. sapiens GALT	Expression of human Gal transferase I.		
D. melanogaster MAN-II	Expression of *Drosophila* mannosidase II.		
T. reesei MAN-I	Expression of *Trichoderma* mannosidase.		

Modification	Description	Output	Reference
Increased productivity			
KAR2	Overexpression of the ER-resident chaperone *KAR2* by P_{AOX1}	High secreting strain producing 3-fold more A33 single- chain antibody over the parent strain	Damasceno et al (2007)
PDI	Overexpression of *S. cerevisiae* protein disulphide isomerase by P_{GAP}.	High secreting strain producing 1.9 fold more 2F5 ant- HIV Fab fragment over the parent strain.	Gasser et al (2005)
AFT1	Overexpression of the *P. pastoris* transcription factor Aft1 involved in the regulation of carbohydrate metabolism and protein secretion.	High secreting strain producing 2.5 fold more Sphingopyxis carboxylesterase than the parent strain.	Ruth et al (2014)
bgs13	Disruption of the *P. pastoris* homologue to *S. cerevisiae* protein kinase C	Mutant strain with a high secretion phenotype – capable of producing ~3 fold more HSA than the parental strain.	Larsen et al (2013)
Protease deficient strains			
Δpep4	Deletion of the gene encoding a major vacuolar aspartyl protease	Protease deficient strains designated SMD1163 (*Δhis4, Δpep4. Δprb1*) and SMD1168 (*Δhis4, Δpep4*). Incorporated into the *Pichia*Pink™ expression platform (Thermo Fisher Scientific)	Gleeson et al (1998)
prb1	Disruption of the gene encoding proteinase B		
Genetic stability of recombinant clones			
Δku70	Deletion of the *P. pastoris KU70* homologue, forming part of the heterodimer initiating non homologous end joining (NHEJ).	NHEJ deficient strain targeted more efficiently by genomic integration vectors. High genetic stability of clones containing 4-7 transgene copies	Näätsaari et al (2012)

1.6 Factors affecting the process efficiency of *P. pastoris*

1.6.1 Bottlenecks in secretion

Despite exhibiting high volumetric productivities, due in part to the high cell densities that can be achieved in fermentor culture, the specific productivity of individual *P. pastoris* cells, relative to other expression systems, has been historically reported as low for secreted heterologous proteins (Krainer et al, 2012). This intrinsic flaw in the *Pichia* system was recently supported in a study comparing recombinant protein secretion between *P. pastoris* and CHO cells, finding that the latter secreted 1011-fold more 3D6 single chain Fv-Fc anti-HIV-1 antibody and 26-fold more human serum albumin (HSA) per unit biomass (Maccani et al, 2014). Different independent studies have shown that a significant proportion of highly expressed heterologous proteins targeted to the secretory pathway accumulate within the endoplasmic reticulum (ER) and Golgi body in an insoluble form, indicating that a limiting factor to the specific productivity of *P. pastoris* is the capacity of its secretory pathway (Inan et al, 2006; Love et al, 2012). Furthermore the activation of cellular processes acting to clear excess protein from the bottlenecked regions of the secretory pathway introduce a number of product loss points throughout the route of protein export (Puxbaum, Mattanovich and Gasser, 2015); the locations of which are illustrated in Figure 4.

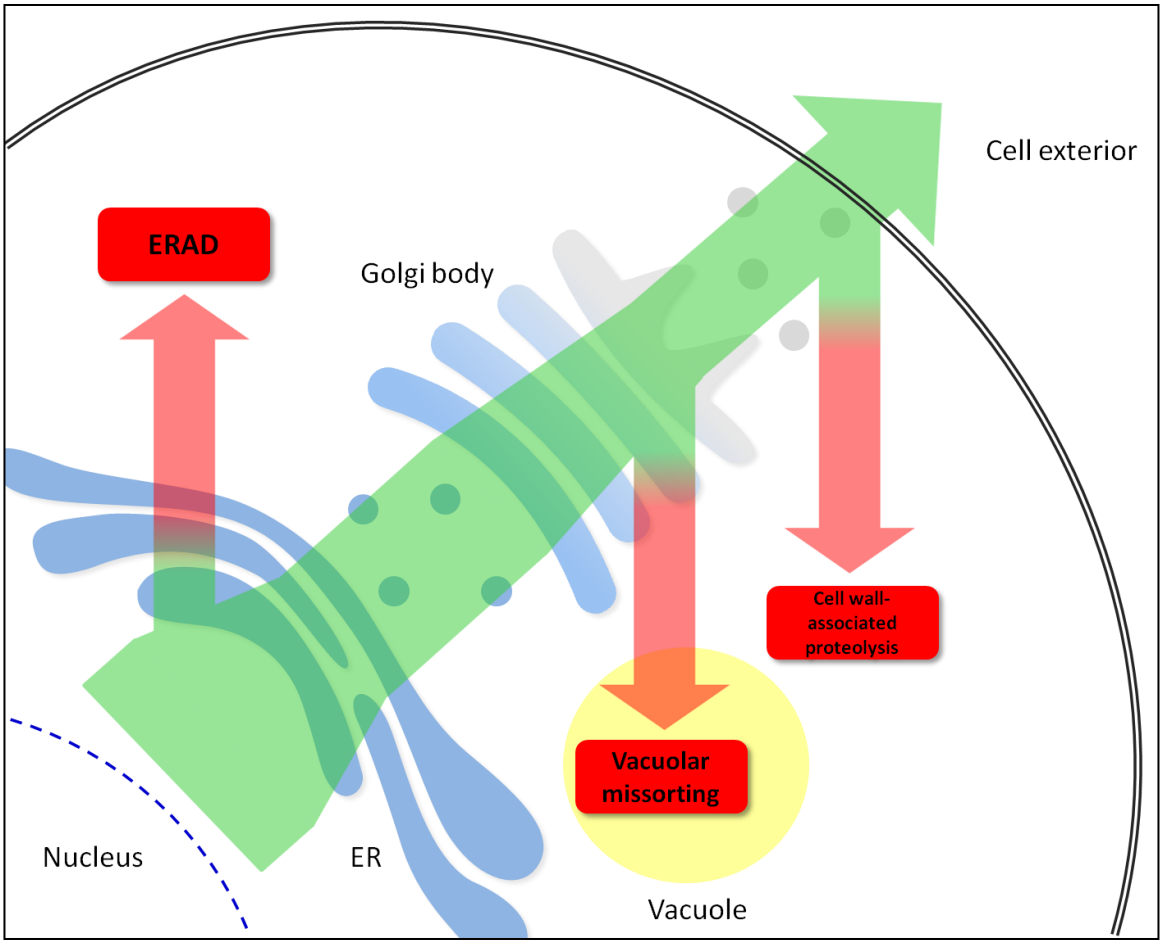


Figure 4: Location of bottlenecks and processes causing loss (red) in the overall titre of heterologous proteins (green) during transport through the secretory pathway in *P. pastoris*.

YPS1, encoding yapsin 1 – a cell wall associated protease involved in cell wall assembly, was found to be involved in the proteolysis of heterologous proteins during exocytosis at the cell periphery (Silva et al, 2011). Deletion of *YPS1* resulted in the increased secretion of collagen inspired proteins as well as an HSA and human parathyroid hormone fusion protein for a double deletion of *YPS1* and *PEP4* (Silva et al, 2011; Wu et al, 2013). Another known bottleneck of heterologous protein secretion is caused by the mistargeting of proteins from the Golgi body to the vacuole through the carboxypeptidase Y (CPY) pathway, where they are subsequently degraded by vacuolar proteases; a process which has been studied more comprehensively for other yeast expression systems (Idiris et al, 2006). Recently though a novel mutant strain of *P. pastoris* with a disrupted vesicular transport pathway between the Golgi and vacuole, coupled with the deletion of a vacuole-resident protease was presented and reported to exhibit increased production of secreted HSA (Gasser, 2016).

However a major bottleneck in secretion that has yet to be addressed fully occurs in the ER and relates to the activation of the ER associated stress pathways: the unfolded protein response (UPR) and endoplasmic reticulum associated degradation (ERAD) in response to its saturation with misfolded heterologous protein (Puxbaum, Mattanovich and Gasser, 2015). This has led to the establishment of the ER in *P. pastoris* and other expression systems as an engineering "hotspot" for research, for which numerous strategies have been developed to artificially increase its capacity for protein processing (Delic et al, 2014).

1.6.1.1 The unfolded protein response (UPR)

The overexpression of recombinant proteins that are foreign to the folding and transport environment of the ER can burden the secretory system by exceeding the ER's folding capacity, leading to the accumulation of insoluble aggregates by the interaction of misfolded proteins (Patil and Walter, 2001). Prolonged ER stress leads to the activation of the unfolded protein response (UPR); a stress response conserved in all eukaryotes that serves to temporarily increase the folding capacity of the ER (Patil and Walter, 2001). The model for the activation of the UPR in yeast has been well studied and, although certain elements of the pathway are still disputed (Pincus et al, 2010), an established model is summarised in Figure 5.

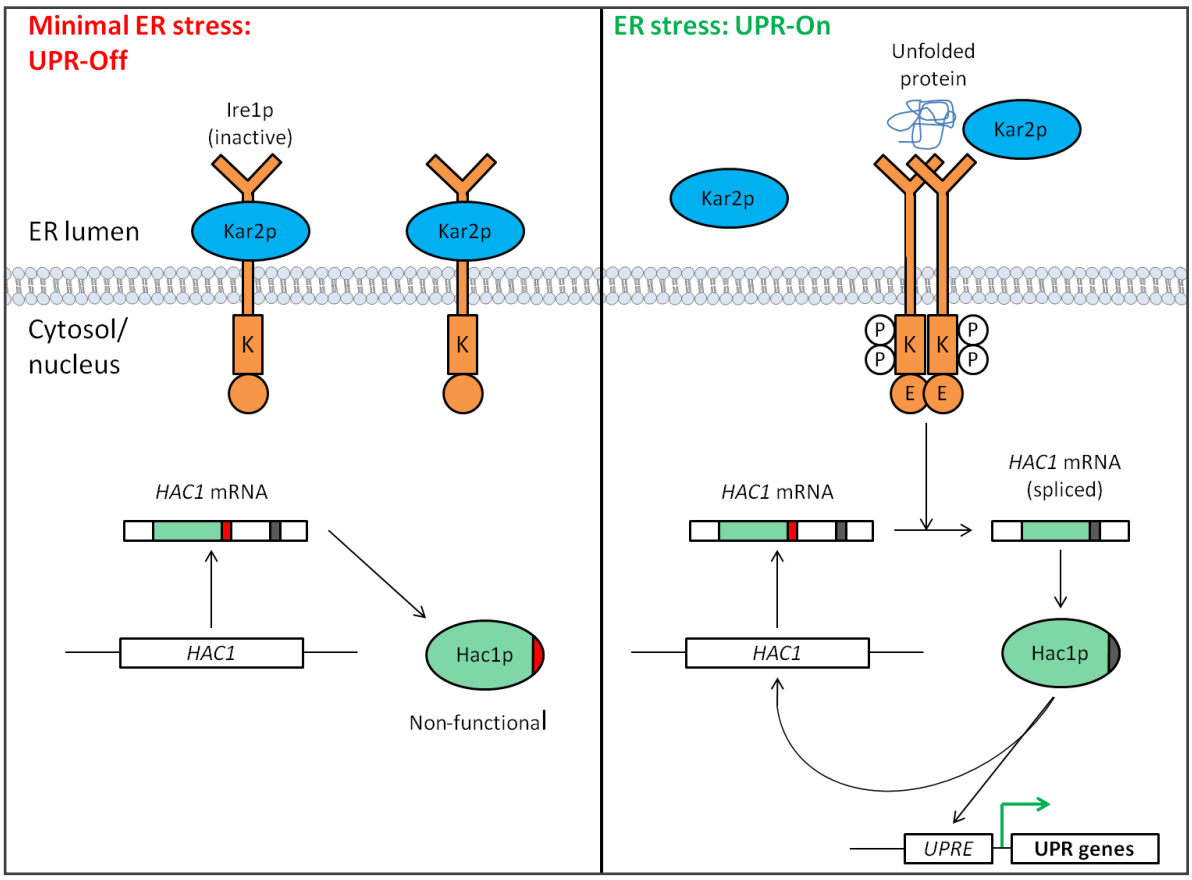


Figure 5: Schematic of the model for the activation of the UPR in yeast. The dimerization of the transmembrane ER protein Ire1p occurs due to the dissociation of Kar2p in response to the build-up of incorrectly folded proteins in the ER. Subsequent autophosphorylation of the cytosolic kinase domains (K) in the Ire1p dimer results in the activation of its endonuclease domain (E), responsible for the alternative splicing of the *HAC1* to yield the functional form of the Hac1p transcription factor. Hac1p binds to UPREs within the promoters of UPR associated genes to upregulate their transcription.

In yeast, including *P. pastoris*, the detection of misfolded proteins, and therefore the primary step in the activation of the UPR, is facilitated by the interaction between the chaperone Kar2p and the ER membrane-bound protein kinase - Ire1p (Cox, Shamu and Walter, 1993). Under non-stressed conditions Kar2p is associated with Ire1p on the ER lumenal side of the transmembrane protein. The dissociation of Kar2 from Ire1p in response to an increase in misfolded proteins within the ER, as well as the direct binding of misfolded proteins to Ire1p leads to the dimerization of Ire1p and the activation of its cytosolic, endonuclease domain (Shamu and Walter, 1996; Credle et al, 2005). The gene encoding the transcription factor for UPR-regulated genes, named *HAC1*, is expressed constitutively (Patil and Walter, 2001). However, under unstressed conditions The *HAC1* gene

product is present in an inactive form due to the presence of an intron in the *HAC1* mRNA transcript (Cox and Walter, 1996). The endonuclease activity of the activated Ire1p serves to splice the intron from the *HAC1* mRNA to produce the active form of HAC1p (Cox and Walter, 1996; Mattanovich et al, 2004). HAC1p in turn binds to the UPR element (UPRE), a 22bp element found upstream of all UPR regulated genes, and activates the expression of genes involved in the UPR response, including ER resident chaperones, foldases and membrane biogenesis factors (Mattanovich 2004).

Research into the *HAC1* homologue in *P. pastoris* revealed some diversification in its regulation from the canonical yeast model of UPR (Guerfal et al, 2010). The spliced form of *HAC1* encoding the active factor was detected under standard growth conditions, suggesting that a basal level of UPR is constitutively active. Whilst *HAC1* is expressed at consistent levels in yeast, the transcription of both variants of *P. pastoris HAC1* was increased during UPR, indicating that Hac1p upregulates its own transcription in a positive feedback loop fashion (Guerfal et al, 2010).

1.6.1.2 Endoplasmic reticulum associated degradation (ERAD)

While the activation of the UPR may at first seem beneficial to the production of heterologous proteins as it aids in the correct processing of proteins and in returning the ER to homeostasis, its prolonged upregulation eventually leads to the increased activation of the endoplasmic reticulum associated degradation pathway (ERAD) (Travers et al, 2000). The accumulation of misfolded proteins in the ER lumen in yeast triggers the activation of a specific arm of the ERAD pathway, named ERAD-L, through the assembly of complex containing the UPR factor Kar2p and Yos9p that specifically bind misfolded proteins (Ismail and Ng, 2006). The elements comprising ERAD-L function to clear the ER lumen of any terminally misfolded protein aggregates by their translocation into the cytosol. Once in the cytosol the misfolded proteins are subsequently tagged for degradation by proteasomal complexes by ubiquitination (Hampton, 2002). Ultimately the prolonged upregulation of the UPR is detrimental to the production of heterologous proteins, not only through the loss of product caused by the ERAD pathway, but also because a number of apoptotic factors have been identified further downstream of the UPR pathway, triggering apoptosis under continued ER stress (Fribley, Zhang and Kaufman, 2009).

Genes involved in the UPR and ERAD pathways have been found to be upregulated during the production of secreted recombinant proteins in *P. pastoris* in multiple studies, implicating ER associated stress as a limiting factor for productivity (Inan et al, 2006; Gasser et al, 2007A; Love et al, 2010; Whyteside et al, 2011). An analysis of productivity and ER stress activation for *P. pastoris* secreting a library of human lysozyme variants differing in their native state stabilities found an inverse relationship between secreted titre of recombinant protein and increasing instability (Kumita et al, 2006). The secretion of more unstable variants also resulted in higher upregulation of UPR and ERAD factors, leading the authors to conclude that proteins with weaker stabilities saturate the folding capacity of the ER in smaller doses due to a higher propensity to misfold, resulting in higher levels of product loss from a more active ERAD (Whyteside et al, 2011). The relationship between protein stability, the induction of ER stress and the rate of intracellular degradation of recombinant proteins provides some insight as to why maximum titres can often vary greatly between different recombinant proteins produced in *P. pastoris*. Some example comparisons include HSA and *E. coli* phytase which can both be secreted to reported titres of over 10 and 6.4g/l respectively whereas the expression Anti-IL6 receptor antibody fragment, used to treat rheumatoid arthritis, yielded only 30mg/l of final product (Bushell et al, 2003; Chen et al, 2004; Zhiyi et al, 2007).

1.6.2 Clonal variation

Clonal variation, within the context of expression systems, describes a phenomenon in which clones derived from daughter cells raised from homogenous cell lines or following transformation display a wide distribution in recombinant protein productivities beyond the expected levels of biological variation. Clonal variation is a widely reported factor of *P. pastoris* expression and necessitates the excessive screening of several colonies post-transformation to eliminate clones that secrete well below average titres of recombinant proteins, and to identify high secretors (Brankamp et al, 1995; Love et al, 2010; Baumann et al, 2011). A known source of interclonal variation, occurring between transformants cloned with an integrated expression vector, is a variation in transgene copy numbers, as it is predicted that multicopy clones arise at a frequency of 1% within a transformed population (Romanos et al, 1998). However an intrinsic level of unaccounted variation is still observable between *P. pastoris* clones secreting identical proteins at equivalent copy numbers (Love et al, 2010; Aw, 2012). A few studies have attempted to analyse clonal variation at the genetic and

transcriptional level. Genetic fingerprinting of 17 different single copy clones transformed with an expression vector for an unspecified protein revealed the presence of random mutations throughout the genome in comparison to control strains; indicated to have occurred during the transformation process (Viader-Salvadó et al, 2006). The highest producers within the tested sample correlated with having the fewest genetic differences from the host strain, implicating the random genetic mutations observed as contributors to clonal variation, although the source of the mutagenesis was not identified (Viader-Salvadó et al, 2006). A microarray analysis was conducted on a subset of high, medium and low secreting single copy clones of recombinant HSA by Aw (2012) to determine any differences in global gene transcription underlying clonal variation. Significant variation was found not only between but within samples of clones with different secretion phenotypes, although genes comprising the ERAD pathway were uniformly upregulated in high secreting samples, suggesting the variation in the degree of its activation as a contributing factor (Aw, 2012).

There is also evidence to suggest that variation is apparent at the intraclonal level and affects the stability of the productivity phenotype of isolated "high secreting" homogenous cell lines through subsequent generations. A study by Love et al (2010) examined the productivity of different cell lines characterised by their rate of secretion of a human Fc fragment after growth to late exponential phase in liquid culture. The results showed that cultures derived from high secreting cells did not retain the parental phenotype and rather displayed a range of secreted productivities similar to the daughter cells of "low secretors" (Love, 2010). The results highlight a potential transient nature of high secreting phenotypes found through recombinant clone screening, and raise concerns for the maintenance of high productivities for recombinant proteins during expression trials requiring longer incubation times.

Other than the studies discussed, little research has been undertaken to define the extent of clonal variation or search for causative factors within the biology of *P. pastoris* and within the techniques used to engineer its expression.

1.7 Aims and objectives

One of the objectives of this project was to engineer new platform strains displaying an increased specific productivity for a host of diverse and industrially relevant recombinant proteins. The project initially focused on attempting to improve productivity by engineering the ER to increase its processing capacity for heterologous proteins while minimising the activation of ER-stress responses, with scope for exploring other process limiting factors identified within the *Pichia* system.

Secondly, this project attempted to build on the limited research into clonal variation by examining underlying factors within the fundamentals of the *Pichia* expression system that could give rise to variation in specific productivity. Studies were divided into searching for inherent, natural variability within the native *P. pastoris* cell population that could predispose cells to secrete heterologous proteins as well as investigating external sources of variation that could be artificially introduced in well established protocols used in the engineering, growth and expression of the *P. pastoris*. The implications of locating factors responsible for clonal variation range from the design of a more robust expression system with a consistent level of productivity, to the potential of defining a set of characteristics in a clonal population that skew variation towards the pre-selection of highly productive strains.

From these broad aims, a set of objectives encompassed by the project were designed as follows:

- The effect on clonal variation from the use of zeocin, a known mutagen, as a selectable marker was examined.
- Variation in native factors between members of set clonal populations such as AOX1-based transcription, UPR activation, and growth characteristics were compared and correlated with productivity to assess their contribution to clonal variation.
- The gene putatively belonging to the Opi1p yeast transcription factor family was knocked out in *P. pastoris* and the resulting mutant strains were tested for specific productivity and ER stress activation for the production of a range of recombinant proteins.
- Strains exhibiting reduced levels of glucose repression of the *AOX1* promoter were engineered, and were tested for recombinant protein productivity through screens and expression trials with glucose/methanol mixed feeds.

2. Materials and Methods

2.1 Strains and plasmids

2.1.1 *E. coli* strains

Table 2: *E. coli* strains used in this study

Strain	Description	Source	Reference
BioBlue	*rec*A1 ,em>endA1 *gyr*A96 *thi*-1 *hsd*R17(r_k-, m_k+) *sup*E44 relA1 *lac* [F' *pro*AB *lac*I^qZΔM15 Tn*10*(Tet^r)]	Bioline	Bioline catalogue
DH5 alpha	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169, hsdR17($r_K^-m_K^+$), λ^-	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB*$^+$ *Δ*(*lac-proAB*) e14- [F' *tra*D36 *pro*AB^+ *lac*I^q *lac*ZΔM15] hsdR17($r_K^-m_K^+$)	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
E.coli pPICZαB	JM109 (pPICZαB)	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
E. coli pAVE522	BioBlue (pAVE522)	Fujifilm Diosynth Biotechnologies	Unpublished
E. coli pIB2	DH5 alpha (pIB2)	Gift from Benjamin Glick (Addgene plasmid #25451)	Sears et al. (1998)
E. coli pPICzm-T70A	JM109 (pPICZαA, HuL-T70A)	Bryn-Edwards Jones, Imperial College	Unpublished
E. coli pGrzαHSA	BioBlue (pGrzαHSA, Zeo^r, *hsa*)	Rochelle Aw, Imperial College	Aw (2012)
E. coli pAG32	DH5 alpha (pAG32)	Gift from John McCusker (Addgene plasmid #35122)	Goldstein and McCusker (1999)
E. coli p414-TEF1p-Cas9-CYC1t	DH5 alpha (p414-TEF1p-Cas9-CYC1t)	Gift from George Church (Addgene plasmid #43802)	Dicarlo et al (2013)

2.1.2 *P. pastoris* strains

Table 3: *P. pastoris* strains used as part of this study

Strain	Description	Source	Reference
NRRL 11430	Wild type strain	Donated by Fujifilm Diosynth Biotechnologies (Billingham, UK)	Northern Regional Research Laboratories (Peoria, USA)
GS115	Derivative of NRRL 11430, Mut+, *his4*	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
CLD804	NRRL11430 (pAVE522; *BPTI*), subjected to PTVA and screened for high productivity	Fujifilm Diosynth Biotechnologies (Billingham, UK)	Unpublished
CLD819	NRRL11430 (pAVE522; *PI3*), subjected to PTVA and screened for high productivity	Fujifilm Diosynth Biotechnologies (Billingham, UK)	Unpublished
CLD883	NRRL11430 (pAVE522; *PRSS1*), subjected to PTVA and screened for high productivity	Fujifilm Diosynth Biotechnologies (Billingham, UK)	Unpublished
GpαGOxZ	GS115 (pPICZαB; *GOX*)	This study	Unpublished
GpαGOxH	GS115 (pPICZαB; *sh bleΔ*::*HIS4*; *GOX*)	This study	Unpublished
NRRL 11430 *Δopi1*	NRRL 11430 (*opi1Δ*::*Tn903kan*^{r})	This study	Unpublished
CLD804 *Δopi1*	CLD804 (*opi1Δ*::*Tn903kan*^{r})	This study	Unpublished
NpαGOxZ	NRRL 11430 (pPICZαB; *GOX*)	This study	Unpublished
NpαGOxZ *Δopi1*	NRRL 11430 (*opi1Δ*::*Tn903kan*^{r}; pPICZαB; *GOX*)	This study	Unpublished
NAαT70A	NRRL 11430 (pPICZαA; *HUL-T70A*)-single copy	This study	Unpublished
NAαT70A *Δopi1*	NRRL 11430 (*opi1Δ*::*Tn903kan*^{r}; pPICZαA; *HUL-T70A*) – single copy	This study	Unpublished
P. pastoris + PAVECRS	NRRL 11430 (pAVE522; *sh bleΔ*::*Tn903kan*^{r}; *sh ble* (under control of the AOX1 promoter))	This study	Unpublished
P. pastoris + PpCas9	NRRL 11430 (pGrzαHSA; *sh bleΔ*::*hph*; *αmf::HSAΔ::CAS9-sv40*)	This study	Unpublished
NGmit1	NRRL 11430 (pGrzαHSA; *sh bleΔ*::*hph*, *αmf::HSAΔ::MIT1*) – single copy	This study	Unpublished

Strain	Description	Source	Reference
NGmit1 +pAVECRS	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*; *αmf-HSAΔ::MIT1*; pAVE522; *sh bleΔ*::*Tn903kan*^{r}; *sh ble* (under control of the AOX1 promoter)) – single copy	This study	Unpublished
NGmit1-AαT70A	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*; *αmf-HSAΔ::MIT1*; pPICZαA; *HUL-T70A*) – single copy	This study	Unpublished
NGmit1 +pAVECRS-mxr1tsv40	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*; *αmf::HSAΔ::MIT1*; pAVE522; *sh bleΔ*::*Tn903kan*^{r}; *MXR1*(1-1200)-*sv40*; *sh ble* (under control of the AOX1 promoter))) – single copy	This study	Unpublished
NGmm1	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*, *αmf::HSAΔ::MIT1-T2A-MXR1*(1-1200)) – single copy	This study	Unpublished
NGmm1-AαT70A	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*, *αmf::HSAΔ::MIT1-T2A-MXR1*(1-1200); pPICZαA; *HUL-T70A*) – single copy	This study	Unpublished
NGmm1 +pAVECRS	NRRL 11430 (pGrzαHSA; *sh ble*::*hph*; *αmf::HSAΔ::MIT1-T2A-MXR1*(1-1200); pAVE522; *sh bleΔ*::*Tn903kan*^{r}; *sh ble* (under control of the AOX1 promoter)) – single copy integrant	This study	Unpublished

2.2 Media

2.2.1 Lysogeny broth (LB) medium

LB medium was composed of 0.5% (w/v) yeast extract (Melford, UK), 1.0% (w/v) peptone from casein (Merck, UK), 1.0% (w/v) NaCl (Sigma-Aldrich, UK). For LB agar plates agar (Melford, UK) was added to a final concentration of 1.6% (w/v). For ampicillin resistance selection ampicillin (Sigma-Aldrich, UK) was added to a final concentration of 100µg/ml. For selection by zeocin resistance zeocin (Thermo Fisher, UK) was added to a final concentration of 25µg/ml. For selection by kanamycin resistance kanamycin (Thermo Fisher, UK) was added to a final concentration of 30µg/ml. For selection by hygromycin resistance hygromycin b (Roche, UK) was added to a final concentration of 100µg/ml.

2.2.2 Super optimal broth with catabolite repression (SOC) medium

SOC medium was composed of 0.5% (w/v) yeast extract (Melford, UK), 2.0% (w/v) peptone from casein (Merck, UK), 10 mM NaCl (Sigma-Aldrich, UK), 2.5 mM KCl (Sigma-Aldrich, UK), 10 mM $MgCl_2$ (Acros, UK), 10 mM $MgSO_4$ (Fisher, UK) and 20 mM dextrose (glucose) (Merck, UK).

2.2.3 Yeast extract peptone dextrose (YPD) medium

YPD medium was composed of 1% (w/v) yeast extract (Melford, UK), 2% (w/v) peptone from casein (Merck, UK), 2% (w/v) dextrose (glucose) (Merck, UK). For YPD agar plates, agar (Melford, UK) was added to a final concentration of 1.6% (w/v). For selection by zeocin resistance zeocin (Thermo Fisher Scientific, UK) was added to a final concentration of 100µg/ml from a 100mg/ml stock solution. For selection by G418 resistance G418 sulphate (Sigma-Aldrich, UK) was added to a final concentration of 500µg/ml from a 50mg/ml stock solution. For selection by hygromycin b resistance, hygromycin b (Sigma-Aldrich, UK) was added to a final concentration of 200µg/ml from a 50mg/ml stock solution.

2.2.4 Minimal dextrose (MD) medium

MD medium was composed of 1.34% (w/v) yeast nitrogen base (YNB) (Sigma-Aldrich, UK), 4×10^{-5}% (w/v) biotin (Sigma-Aldrich, UK), 2% (w/v) dextrose (glucose) (Merck, UK). For MD agar plates, agar (Melford, UK) was added to a final concentration of 1.6% (w/v).

2.2.5 Minimal methanol (MM) medium

MM medium was composed of 1.34% yeast nitrogen base (YNB) (w/v) (Sigma-Aldrich, UK), 4×10^{-5}% (w/v) biotin (Sigma-Aldrich, UK), 0.5% (v/v) methanol (Sigma-Aldrich, UK). For MM agar plates, agar (Melford, UK) was added to a final concentration of 1.6% (w/v).

2.2.6 Minimal dextrose and methanol (MDM) medium

MDM medium was composed of 1.34% yeast nitrogen base (YNB) (w/v) (Sigma-Aldrich, UK), 4 x 10^{-5}% (w/v) biotin (Sigma-Aldrich, UK), 0.5% (v/v) dextrose (glucose) (Merck, UK), 0.5% (v/v) methanol (Sigma-Aldrich, UK). For MDM agar plates, agar (Melford, UK) was added to a final concentration of 1.6% (w/v).

2.2.7 Buffered minimal glycerol medium (BMG) and buffered minimal methanol medium (BMM)

BMG/BMM medium was composed of 1.34% (w/v) yeast nitrogen base (YNB) (w/v) (Sigma-Aldrich, UK), 4 x 10^{-5}% (w/v) biotin (Sigma-Aldrich, UK), 100mM potassium phosphate pH 6.0 (Fisher, UK), 1% (v/v) glycerol (BDH Prolabo, UK) or 0.5% (v/v) methanol (Sigma-Aldrich).

2.2.8 Buffered glycerol-complex medium (BMGY) and buffered methanol-complex medium (BMMY)

BMGY/BMMY medium was composed of 1% (w/v) yeast extract (Melford, UK), 2% (w/v) peptone from casein (Merck, UK), 1.34% (w/v) yeast nitrogen base (YNB) (w/v) (Sigma-Aldrich, UK), 4 x 10^{-5}% (w/v) biotin (Sigma-Aldrich, UK), 100mM potassium phosphate pH 6.0 (Fisher, UK), 1% (v/v) glycerol (BDH Prolabo, UK) or 0.5% (v/v) methanol (Sigma-Aldrich).

2.3 Molecular biology methods

2.3.1 End point Polymerase chain reaction (PCR)

PCR reactions were carried out in 0.25ml PCR tubes (Fisher, UK) using the Bioer Genepro thermal cycler (Alpha Laboratories, UK). 50µl reactions were composed of: 1µl DNA template, 5µl 2.5mM dNTPs, 5µl 5µM each primer, 10µl 5X Phusion HF buffer or 5X Q5 buffer, 1 unit Phusion Hot Start II polymerase (Thermo Fisher, UK) or Q5® High-Fidelity DNA Polymerase (New England Biolabs, UK) and Milli-Q water (18.2 mΩ). PCR reactions were carried out using the conditions specified in the instructions for the appropriate polymerase for 35 cycles, although the annealing temperature was varied depending on the T_m of the primers used within the reaction.

2.3.2 Diagnostic/ colony PCR

20µl reactions using KAPATaq Ready Mix DNA Polymerase (Anachem, UK) were composed of: 10µl 2x KAPATaq ReadyMix with Mg^+, 1.6µl each primer (5µM), 1µl genomic DNA template for *P. pastoris* or bacterial colony added directly in the case of *E. coli*, and Milli-Q water (18.2 mΩ). PCR reactions were carried out under the conditions stated in the user manual for 35 cycles, although the annealing temperature was varied depending on the T_m of the primers used within the reaction. The time of initial denaturation was set to 5 minutes for colony PCRs.

2.3.3 Quantitative Polymerase chain reaction (qPCR)

qPCR was implemented in this project to primarily determine recombinant gene copy numbers in *P. pastoris*. Templates of genomic DNA were prepared using the YeaStar™ Genomic DNA kit (Cambridge Bioscience, UK), quantified and diluted to a working stock concentration of 10ng/µl.

qPCR primers for target genes were designed using the online tool Primer3 (http://bioinfo.ut.ee/primer3/) with the following constraints:

- The amplified region of DNA does not exceed 200bp.
- Primer size ranges from 18-24bp.
- GC content of 50-60%.
- Tm range between 60-63°C.
- Maximum Tm difference of 5°C (although 1°C was optimal)
- Maximum 3' self complementary score (achieved by global alignment of 3' regions of primer pairs to predict the likelihood of primer dimer formation) of 1.
- Maximum size of mononucleotide repeats restricted to 3.

3µl of working stock DNA template was loaded onto 96 – well PCR plates (Thermo Fisher Scientific, UK). Forward/reverse qPCR primers and 2X LuminoCt® SYBR® Green qPCR ReadyMix™ (Sigma-Aldrich, UK) were added to final concentrations of 0.25µM each and 1X respectively within a final reaction volume of 20µl after which the samples were centrifuged at 2000rpm for 1 minute before running. qPCR reactions were run in triplicate for each sample, using the DNA Engine® Peltier thermal cycler coupled with the Chromo4™ Real-Time PCR Detector (BioRad, Hemel Hempstead, UK) and data was outputted through the Opticon 3 thermal cycler software program (BioRad, Hemel Hempstead, UK). qPCR conditions were set as specified by the manufacturer's instructions for the qPCR of genomic templates with SYBR® Green qPCR ReadyMix™, specifically an initial denaturation step of 95°C for 2 minutes, followed by 40 cycles consisting of a denaturation step of 95°C for 3 seconds and an annealing/extension step of 60°C for 30 seconds. A melt curve to confirm the formation of single products during qPCR was performed for each sample, in which the temperature was raised from 55°C to 95°C and fluorescence readings were taken at 0.2°C increments.

Cycle threshold (C_t) values obtained for each sample were interpreted using the method first described by Pfaffl (Pfaffl, 2001) and applied by Abad et al (2010) to determine gene copy numbers in *P. pastoris*. The formula is given as:

$$R = \frac{(E_{target\ gene})^{\Delta Ct\,target\ (control-sample)}}{(E_{reference\ gene})^{\Delta Ct\,reference\ (control-sample)}}$$

In which "R" represents the relative abundance of the target gene to the reference gene, "E" is the efficiency of amplification achieved through 1 cycle by the specific primer pair amplifying the detected region of either the target or reference gene and ΔC_t is calculated as the difference in C_t values between the no template control/ negative control and the sample. The gene encoding beta-actin (*ACT1*) in *P. pastoris* was selected as a reference gene for copy number determination due to its presence in the genome as a single copy.

2.3.4 Storage of *P. pastoris* cell samples for RNA isolation

For cell samples that were not processed immediately for RNA purification after growth, 600µl of culture was diluted in 3ml Ambion RNAlater® solution (Thermo Fisher Scientific) and incubated at room temperature for 1 hour. Cell suspensions were then centrifuged at 4000 rpm for 5 minutes and resuspended in 1ml RNAlater®. Samples were stored at -80°C prior to RNA purification.

2.3.5 Isolation of RNA from *P. pastoris* cells

RNA was isolated from cell samples of *P. pastoris* using the GenElute™ Total RNA Purification Kit (Sigma Aldrich, UK) according to the manufacturer's instructions for the preparation of RNA from yeast samples, with some modifications to improve cell lysis and the removal of genomic DNA. During the cell wall digestion step lysozyme concentration was increased from 1U/µl to 2U/µl and the digest incubation step was extended to 1.5 hours. During cell lysis the vortexing of samples containing lysis buffer RL was extended to 1 minute and the incubation period during the removal of genomic DNA with the On-Column DNase I Digestion Set (Sigma Aldrich, UK) was extended from 15 minutes to 45 minutes. Prior to elution the eluting solution was warmed to 60°C to increase the final RNA yield. RNA concentration and purity was checked with the NanoVue Plus spectrophotometer (GE Healthcare, UK).

2.3.6 Reverse transcription of RNA samples

The Applied Biosystems High Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, UK) was used for the preparation of cDNA from purified RNA samples according to the manufacturer's instructions. 1µg of RNA sample diluted in 10µl nuclease – free water (Thermo Fisher Scientific) was used in a single reaction. cDNA samples were subsequently stored at -20°C.

2.3.7 Reverse Transcription Quantitative Polymerase chain reaction (RT-qPCR)

The protocol and data analysis for RT-qPCR, to identify changes in gene transcription as part of this project, is similar to the methods previously described for qPCR with the exception of the preparation of template for samples and the conditions of the PCR reaction. Following reverse transcription of isolated RNA samples, a 1 in 10 dilution of the resulting cDNA in Milli-Q water (18.2 mΩ) was prepared to give a working stock solution at a concentration of 5ng/µl based on the concentration of the original RNA sample. 5µl of diluted template was used per 20µl reaction and the initial denaturation step of the PCR was shortened from 2 minutes to 20 seconds.

2.3.7.1 Determination of amplification efficiency for qPCR primer sets

To improve the validity of results obtained from qPCR and RT-qPCR experiments, 3-5 qPCR primer sets for a single target gene were tested for amplification efficiency prior to their use, unless previously validated. A 10 – fold serial dilution of pure genomic DNA or cDNA was carried out to provide a range of 5 template quantities of 0.001ng, 0.1ng, 1ng, 10ng and 100ng per 20µl qPCR reaction. qPCR with the tested primer pairs was conducted in triplicate for each condition and the final C_t values were plotted against the logarithm of the original template quantity. The gradient of the linear plot produced could be used to calculate the efficiency for the primer set using the formula:

$$E = 10^{(\frac{-1}{g})} \times 100\%$$

Where "E" is the amplification efficiency (%) and "g" is the gradient. Primer sets that displayed the highest efficiencies and only formed single products during PCR, as determined by melt curve analysis or gel electrophoresis samples if necessary, were used in qPCR experiments. qPCR primer sets used in this project are displayed in Table 4, and full sequences can be found in 9.1.

Table 4: List of primers used in qPCR and RT-qPCR experiments with their associated efficiencies.

Gene	Primer set	Amplification efficiency/ %
ACT1	30-qACT1-F 31-qACT1-R	93.74
GOX	40-qGOX2-F 41-qGOX2-R	91.35
sh ble	93-qshble2-F 94-qshble2-R	93.02
hph	114-qhygb2-F 115-qhygb2-R	89.18
HULT70A	83-qsynHuL3-F 84-qsynHuL3-R	99.26
HAC1	32-qHAC1-F 33-qHAC1-R	89.24
PDI	36-qPDI-F 37-qPDI-R	89.84
KAR2	34-qKAR2-F 35-qKAR2-R	89.04
MIT1	116-qmit1-F 117-qmit1-R	97.22

2.3.8 Gel electrophoresis of DNA

For the separation of DNA fragments within a sample by size, 0.8% agarose (Sigma-Aldrich, UK) gels were used. Agarose gels were set in 1X TAE buffer, diluted from a 50X stock solution containing

24.2% (w/v) Tris base, 5.71% (v/v) acetic acid, 10% (v/v) 0.5 M EDTA, pH 8.0. 10,000X SYBR® Safe (Thermo Fisher, UK) was added to the agarose gel after heating and before setting, to a final relative concentration of 0.5X. The DNA molecular weight marker used in this study was the GeneRuler™ 1kb DNA Ladder (Figure 6) (Thermo Fisher Scientific, UK), of which 5µl aliquots were loaded in a well for each DNA gel. 5X DNA loading buffer (30% [v/v] glycerol, 0.25% [w/v] bromophenol blue) was added to samples to a 1X final relative concentration. Agarose gels were normally run in Bio-Rad Wide Mini-Sub Cell GT Tanks (Bio-Rad, UK) containing 1X TAE buffer at 100V for approximately 40 minutes. The Bio-Rad Power Pac Basic (Bio-Rad, UK) was used as a power source. A Syngene G:Box ChemiHR system (Syngene, UK) was used to visualise and image DNA in agarose gels through a short wave UV transilluminator.

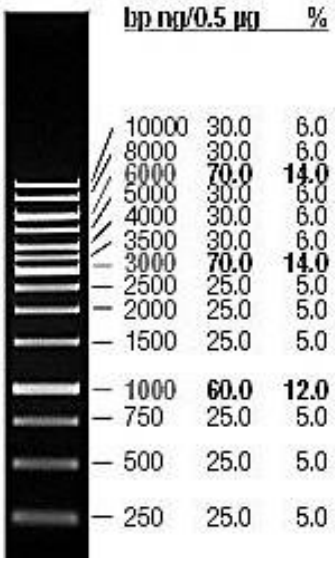


Figure 6: Estimation of DNA fragment sizes from the GeneRuler™ 1kb DNA Ladder used in this project.

2.3.9 Purification of DNA from agarose gels

DNA in agarose gels was visualised by exposure to UV light from a UV box (Fotodyne, USA), and bands were cut out with a scalpel. The Zymoclean™ Gel DNA Recovery Kit (Cambridge Bioscience, UK) was used to purify DNA from the gel blocks, cut out from the agarose gels, according to the manufacturer's instructions.

2.3.10 Purification and concentration of DNA in solution.

2 different methods were used to purify and concentrate DNA. The DNA Clean & Concentrator -5 kit from Zymo research (Cambridge Bioscience, UK) was typically used to concentrate plasmid DNA from miniprepped samples into a volume of 10µl in Milli-Q water (18.2 mΩ).

Ethanol precipitation was used when a more efficient method for concentrating higher quantities of DNA was required, such as for the preparation of linearized plasmid for transformation into *P. pastoris*. For every 10µl of DNA sample 1µl 3M sodium acetate buffer pH 5.2 (Thermo Fisher Scientific, UK), 20µl 100% ethanol (Sigma Aldrich, UK) and 0.5µl 0.6% (w/v) blue dextran solution (Sigma Aldrich, UK) were added. Samples were then incubated at -20°C for 1 hour to overnight depending on the total DNA quantity within the sample, and were then centrifuged at 14000rpm, at 4°C for 1 hour. The resulting DNA pellets were washed twice by adding 150µl 70% ethanol without resuspending the pellet, centrifuging at 4°C for 15 minutes and air drying at 37°C. The pellets were then typically resuspended in 10µl Milli-Q water (18.2 mΩ).

2.3.11 DNA/ RNA quantification

DNA and RNA concentration was measured using the NanoVue Plus spectrophotometer (GE Healthcare, UK). 3µl of DNA sample was loaded and quantified by measuring the absorbance of light at 260nm. Purity of the sample and the presence of protein contaminants in the sample were quantified by calculating the ratio of absorbance at 260nm to 280nm.

2.3.12 Plasmid purification from *E. coli* cultures

Plasmid DNA was purified from *E. coli* cultures, grown overnight in 5ml LB media, using the Qiaprep® Spin Miniprep Kit (Qiagen, UK) according to the manufacturer's instructions.

2.3.13 DNA restriction

Restriction enzymes were purchased from Thermo Fisher Scientific (UK) or New England Biolabs (UK), and restriction digests were carried out according to the manufacturer's instructions. Most regular DNA restriction reactions were carried out in 20µl volumes at 37°C for 3-4 hours, or 1 – 2.5 hours for fast digest enzymes. Double digests were performed using a compatible buffer for both enzymes and, if the specific buffer was not available, sequential digests were used. Restriction products were run on DNA gels and subsequently gel purified.

2.3.14 DNA ligation

T4 DNA ligase (Thermo Fisher Scientific, UK) was used for DNA ligations in 10µl volumes according to the manufacturer's instructions. An insert to vector ratio of 3:1 was used and ligations were incubated at room temperature for 30 minutes before transformation into chemically competent *E. coli*.

2.3.15 Gibson assembly

An adaptation of the original Gibson assembly protocol (Gibson, 2009) was implemented for the assembly of novel vectors without the need to PCR amplify the vector backbone. PCR primers for amplifying the insert were designed to include ~40bp linkers sharing identity with the regions flanking the site of insertion in the vector, at both the 5' and 3' ends of the final PCR product. A master mix containing 320µl 5x isothermal reaction buffer (25% PEG-8000, 500 mM Tris-HCL pH7.5, 50 mM $MgCl_2$, 50 mM DTT, 1mM full set of dNTPs and 5 mM NAD), 0.64 µl 10U μL^{-1} T5 exonuclease (New England Biolabs, UK), 20 µl 2U μL^{-1} Phusion DNA polymerase (New England Biolabs, UK), 160 µl 40U μL^{-1} Taq DNA Ligase (New England Biolabs, UK) and Milli-Q water (18.2 mΩ) to a final volume of 1.2ml. 15µl aliquots of master mix were then prepared from the total volume and stored at -20°C. 5µl volumes containing 100ng vector, linearized at the site of insertion, and an equimolar quantity of insert were then added to a single aliquot of the master mix and incubated for 1 hour at 50°C in the

Bioer Gene Pro thermal cycler (Eppendorf, Stevenage, UK). 5µl of the assembled sample was then transformed into chemically competent *E. coli* or PCR amplified.

For Gibson assembly reactions between PCR amplified parts ~20bp linker regions were added to the 5' and 3' ends of each PCR product as standard.

2.3.16 Transformation of chemically competent *E.coli* by heat shock

A 20µl aliquot of chemically competent *E. coli* BioBlue cells was thawed on ice for approximately 5 minutes. 5-10µl of the ligation or Gibson assembly mix was added to the cells, which were then incubated on ice for 20 minutes. The cells were then heat shocked at 42°C for 45 seconds, and then held on ice for 2 minutes. The cells were recovered in 1ml SOC medium at 37°C for 1 hour with shaking at 250rpm. Following recovery 100µl of the transformed cell sample was plated onto a suitable LB + antibiotic plate and incubated at 37°C overnight.

2.3.17 Transformation of *P. pastoris* by electroporation

For the targeted integration of expression vectors into the *P. pastoris* genome 5-10µg of vector was linearized by a single digest within the vector's region of homology to the relevant locus within the *P. pastoris* genome (e.g. the *AOX1* promoter region in pPICZ based vectors) and ethanol precipitated into a 10µl final volume.

The protocol for the preparation of competent *P. pastoris* cells was adapted from the EasySelect™ *Pichia* Expression Kit (Thermo Fisher, UK) instruction to be more suited to a smaller sample size. 10ml of YPD was inoculated with *P. pastoris* from a -80°C stock or colony and incubated at 30°C with shaking at 250rpm for ~16 hours. The culture was then diluted into a final volume of 100ml YPD in a 500ml flask to an OD_{600} of ~0.8. The culture was grown aerobically at 30°C at 250rpm until an OD_{600} of 1.3-1.5 had been reached, after which it was decanted into 2 sterile 50ml Centrifuge tubes, and transferred to ice. During the wash stage the cells were centrifuged at 4000rpm for 5 minutes at 4°C, and resuspended in 50ml ice cold, sterile dH_2O. The centrifugation step was repeated 3 more times,

resuspending in 25ml ice cold dH2O, 2ml ice cold 1M sorbitol (Sigma-Aldrich, UK) and finally 400µl ice cold 1M sorbitol respectively. Competent *P. pastoris* cells prepared with this method exhibit a decline in transformation efficiency once stored so fresh competent cells were used for every transformation.

During the transformation step an 80µl aliquot of competent cells, prepared in the previous step, was added to each linearized vector sample, transferred to 2mm electroporation cuvettes (Bio-Rad, UK) and kept on ice for 5 minutes. The GenePulser electroporater (Bio-Rad, UK) was used to pulse the cells at 2000V, 25 µF, 200 Ω, for approximately 5 milliseconds. Immediately following electroporation 1ml ice cold 1M sorbitol was added to each sample and the cells were recovered by incubating at 30°C without shaking for 2 hours. Once the recovery stage was complete 100µl of the transformed cell samples were spread onto the appropriate selective agar plate, and incubated at 30°C for 3 to 5 days.

For experiments requiring a higher efficiency of transformation an electroporation protocol involving the pretreatment of cells with lithium acetate and dithiothreitol (DTT), developed by Wu and Letchworth (2004) was used. Initial growth of *P. pastoris* cells was conducted in the same manner as the in the standard electroporation protocol until, once cells had reached and OD_{600} of 1.3, 11ml samples of culture, equating to approximately $8.0x10^8$ cells, were centrifuged at 4000 rpm for 5 minutes, resuspended in 8ml of 100mM lithium acetate (Sigma Aldrich, UK), 10mM DTT (Thermo Fisher Scientific, UK), 0.6M sorbitol and 10mM Tris-HCl pH 7.5 (Thermo Fisher Scientific, UK) and incubated at room temperature for 30 minutes. The cell suspensions were then pelleted by centrifugation at 4000rpm for 5 minutes and resuspended 1.5ml ice cold 1M sorbitol. 2 wash steps consisting of centrifuging and resuspending cells in 1.5ml ice cold 1M sorbitol were then carried out before resuspending cells in a final volume of 80µl ice cold 1M sorbitol as single aliquots for transformation. Electroporation of cells was then conducted as standard immediately following the preparation of electrocompetent cells.

2.3.18 Isolation of *P. pastoris* genomic DNA

A simple method for the extraction and purification of genomic DNA from yeast cells, developed by Lõoke, Kristjuhan and Kristjuhan (2011) was used to prepare genomic DNA from *P. pastoris* colonies for end-point PCR applications. Single colonies of *P. pastoris* were suspended in 100µl of solution containing 200mM lithium acetate and 1% SDS (Thermo Fisher Scientific), mixed by vortexing and incubated at 70°C for 5 minutes on a heat block. 300µl ice cold 96%-100% ethanol was added after which the samples were vortexed and centrifuged at 15,000g for 3minutes to pellet precipitated DNA. The supernatant was removed by aspiration and samples were washed by resuspending in 500µl ice cold 70% ethanol and centrifuging 15,000g for 3minutes. The supernatant was removed and samples resuspended in 100µl Milli-Q water (18.2 mΩ). Cell debris was removed by centrifuging the sample at 15,000g for 1 minute. The resulting supernatant containing dissolved genomic DNA was removed and stored at -20°C.

For applications requiring samples with higher purity, such as qPCR – experiments, genomic DNA was isolated from *P. pastoris* colonies using the YeaStar Genomic DNA kit from Zymo Research (Cambridge bioscience, UK) according to the manufacturer's instructions.

2.4 Cell biology methods

2.4.1 Storage of *E. coli* and *P. pastoris* strains

Single *E.coli* or *P. pastoris* colonies were inoculated in 3ml LB at 37°C or YPD at 30°C respectively, and incubated at 250rpm overnight. Overnight *E.coli* cultures were stored in 20% glycerol (BDH Prolabo, UK) in sterile 2ml cryogenic vials (Thermo Fisher Scientific, UK) at -80°C whereas *P. pastoris* strains were stored in 15% glycerol at -80°C.

2.4.2 Raising clonal populations of *P. pastoris*

5ml YPD was inoculated with a single colony of the desired *P. pastoris* strain, or a -80°C stock originating from a single colony, and grown overnight for ~16 hours at 30°C, 250rpm. The culture was serially diluted 10 fold in phosphate - buffered saline solution (PBS), consisting of 137mM NaCl (Sigma Aldrich, UK), 2.7mM KCl (Thermo Fisher Scientific, UK), 10mM Na_2HPO_4 (Thermo Fisher Scientific, UK), 1.8mM KH_2PO_4 (Thermo Fisher Scientific, UK), with pH adjusted to 7.4 with HCl, to give a range of diluted samples with dilution factors up to 10^6 of the original culture. 100µl of the dilutions were plated onto YPD plates to raise control populations exhibiting a native background of variation, and on the appropriate selective plates for the tested recombinant strain to raise populations exposed to a selective background during colony growth. Plates containing adequately spaced, discrete colonies were kept and specified samples of colonies raised in each condition were selected at random as a representation of the overall clonal population to progress with expression experiments.

2.4.3 Preparation and growth of *P. pastoris* strains on zeocin gradient agar plates

In order to provide a semi quantitative measure for the activation state of the AOX1 promoter in response to growth on different carbon sources, strains expressing the zeocin resistance marker gene sh ble under P_{AOX1} were assayed for viability on plates containing a concentration gradient of zeocin.

Gradient plates were set up by elevating 1 side of a Sterilin 100mm square Petri dish (Thermo Fisher Scientific, UK) by approximately 6mm, through resting on the end of a 20µl pipette tip (Anachem, UK). Molten 1.6% minimal medium agar containing the appropriate growth substrate was poured into each plate and allowed to set, forming a uniform agar slant across the plate. Agar slants were overlayed with molten minimal medium agar matching the growth substrate composition of the slant, with the addition of 1000µg/ml zeocin (Thermo Fisher Scientific, UK). Approximately 1cm of the top portion of the slant was not overlayed to create a control area on the plate containing ~0µg/ml zeocin. Zeocin gradient plates were stored overnight at 5°C to allow the downward diffusion of zeocin through the slant to create a gradient of zeocin concentration ranging from ~0µg/ml - 1000µg/ml across the plate. Prepared gradient plates were not stored for longer than 24 hours prior

to use to prevent further diffusion of zeocin through the agar, resulting in loss of the concentration gradient.

P. pastoris strains were picked from single colonies on agar plates, inoculated into 5ml minimal media containing the desired carbon source and incubated at 30°C at 200rpm for a maximum of 16 hours. OD_{600} was measured and cultures were normalised to a final OD_{600} of 0.1 in PBS. Using a grid as a template, normalised cell suspensions were applied to gradient plates matching the composition of the liquid media used for growth in 4µl droplets across the zeocin concentration gradient. Droplets were allowed to dry and the plates were incubated at 30°C for 3-4 days depending on the relative growth rates of cells on the growth substrate. Plates were imaged in white light from the Syngene G:Box ChemiHR system (Syngene, UK) and the paired GeneSNAP software (Syngene, UK).

2.4.4 Secreted recombinant protein expression from *P. pastoris*

Small scale expression studies in *P. pastoris* were conducted in 24 deep well micro titre plates (Whatman, UK), 50ml centrifuge tubes (Thermo Fisher Scientific), or in 250ml baffled cell culture flasks (Thermo Fisher Scientific, UK) using protocols adapted from the EasySelect™ *Pichia* Expression Kit instruction manual (Thermo Fisher Scientific, UK)

Single colonies or -80°C stocks of *P. pastoris* were inoculated into 5ml BMGY or BMG in 50ml centrifuge tubes for aerobic growth and incubated at 30°C at 250rpm for ~16 hours. The OD_{600} of each cell culture was measured, and used to calculate the volume of starter culture required to give a final OD_{600} of 1-5 in 3ml of media, depending on the recombinant protein being expressed. The required volume for each cell culture was then centrifuged at 4000rpm for 5 minutes, and the cell pellets were resuspended in 3ml of BMMY or BMM unless stated otherwise. The cells were incubated in 24 deep well plates (Whatman, UK) with gas permeable seals (Thermo Fisher Scientific) at 28°C, 200rpm for 48 hours to 5 days depending on the recombinant protein expressed, and expression was maintained by adding 100% (v/v) methanol to a final concentration of 0.5% (v/v) every 24 hours.

Expression in 50 ml Centrifuge tubes followed the same protocol as expression in 24 deep well plates, except that 5ml volumes were used throughout expression as opposed to 3ml. To improve aerobic

conditions throughout growth and expression the lid of each Centrifuge tube was loosened and secured with autoclave tape.

For expression in 250ml baffled flasks, cultures grown in 5-10ml BMGY/BMG for ~16 hours prior to expression were normalised to the appropriate OD_{600} in 25ml BMMY and incubated in flasks with gas permeable seals.

2.4.5 $DiOC_6(3)$ staining and confocal microscopy

$DiOC_6(3)$ was obtained as part of the Yeast Mitochondrial Stain Sampler Kit from Thermo Fisher Scientific (UK) and diluted to a working stock concentration of 1mg/ml in 100% (v/v) ethanol. 5ml of MD medium was inoculated with *P. pastoris* cells from colonies grown on agar plates and grown overnight at 30°C/ 250rpm. Before staining, individual 22x22mm coverslips (Thermo Fisher Scientific, UK) were placed into 6-well tissue culture plates (Corning, UK) and coated in 500µl 0.01% Poly-L-lysine (mol wt 70,000-150,000) (Sigma Aldrich, UK) for 10 minutes at room temperature. Excess poly-L-lysine was removed by washing the wells 3 times with 1ml PBS. Cells were resuspended to a final OD_{600} of 2.0 in 10mM HEPES (Sigma Aldrich, UK) buffer pH 7.4 with 5% (w/v) glucose and 100µl of suspensions were plated onto poly-L-lysine coated coverslips, held for 5 minutes and washed 3 times with 1ml PBS. Adhered cells were then fixed with 500µl/well 4% (w/v) paraformaldehyde (Sigma Aldrich, UK) for 30 minutes and washed 3 times with 1ml PBS. The $DiOC_6(3)$ working stock solution was diluted to a final concentration of either 0.5, 0.75 or 1.0µg/ml in 10mM HEPES buffer pH 7.4 and used to stain adhered cells by adding 300µl/well to coverslips, incubating for 30 minutes at room temperature and washing 5 times with 1ml PBS. Coverslips were then mounted onto microscope slides (VWR, UK) with VectaShield Antifade Mounting Medium (H-1000) (Vector Laboratories, UK) as the mounting media and samples were immediately imaged with the LSM Meta 510 confocal microscope (Zeiss, UK). Cells were viewed at x63 optical zoom with excitation at 488nm.

2.5 Protein methods

2.5.1 SDS-PAGE of protein samples

2.5.1.1 Preparing SDS-PAGE gels

SDS-PAGE gels were assembled between glass plates in a casting frame (BioRad, UK) using dH_2O, a 30% (w/v) acrylamide mix from National Diagnostics (Thermo Fisher Scientific, UK), 1.5M Tris-Cl, pH 8.8 for resolving gels, 1.0M Tris-Cl, pH 6.8 for stacking gels, SDS (Thermo Fisher Scientific, UK), ammonium persulphate (Thermo Fisher Scientific, UK) and Tetramethylethylenediamine (TEMED) from National Diagnostics (Thermo Fisher Scientific, UK). Volumes of each component used to make each gel type used in this study are listed in Table 5.

Table 5: List of components used to make the resolving and stacking constituents of SDS-PAGE gels.

Component	Volume required for a 5ml total gel volume/ ml		
	12% resolving gel	**15% resolving gel**	**Stacking gel**
dH_2O	1.6	1.1	3.4
30% (w/v) acrylamide mix	2.0	2.5	0.85
1.5M Tris-Cl, pH 8.8	1.3	1.3	0.0
1.0M Tris-Cl, pH 6.8	0.0	0.0	0.65
10% (w/v) SDS	0.05	0.05	0.05
10% (w/v) ammonium persulphate	0.05	0.05	0.05
TEMED	0.002	0.002	0.002

2.5.1.2 Sample preparation and running conditions

5X SDS-PAGE loading buffer, made up of 10% (w/v) SDS, 20% (v/v) glycerol, 10mM DTT, 0.2M Tris (pH 6.8) and 0.05% Bromophenol blue (Sigma Aldrich, UK) was added to culture supernatant samples to a final concentration of 1X. Samples were then incubated at 90°C for 5 minutes to denature proteins

before loading. 20µl of prepared samples were loaded onto SDS-PAGE gels with either 6µl Unstained Protein Molecular Weight Marker (Thermo Fisher Scientific, UK) or PageRuler™ Unstained Low Range Protein Ladder (Thermo Fisher Scientific, UK), and a protein standard (Sigma Aldrich, UK) of a defined concentration prepared under the same conditions.

SDS-PAGE gels were run in SDS-PAGE running buffer (25mM Tris base, 192mM glycine and 0.1% SDS) at 25mA per gel for 1 hour, or until the dye front had migrated to the bottom of the gel, using the Mini-PROTEAN® Tetra System (BioRad, UK) connected to the PowerPac™ Basic (BioRad, UK)

2.5.1.3 Staining/ Destaining

Following electrophoresis SDS-PAGE gels were washed in dH_2O and stained with Coomassie Blue stain solution (0.2% (w/v) Coomassie blue (Thermo Fisher Scientific, UK), 10% (v/v) acetic acid (Thermo Fisher Scientific, UK), 50% (v/v) methanol) for 2 hours. Gels were then washed in dH_2O to remove excess stain and then de-stained in 10% (v/v) acetic acid, 30% (v/v) methanol until protein bands could be visualized clearly with minimal background.

For SDS-PAGE experiments in chapter 3 SimplyBlue™ SafeStain (Thermo Fisher Scientific, UK) was used to stain gels. Gels were stained in SimplyBlue™ SafeStain for 1 hour, washed in dH_2O and destained overnight in dH_2O.

2.5.1.4 Visualisation and image analysis

SDS-PAGE gels were visualised in white light from the Syngene G:Box ChemiHR system (Syngene, UK) and the paired GeneSNAP software (Syngene, UK) or the GS-800™ Calibrated Densitometer (BioRad, UK). Relative quantification of protein quantities in each sample by densitometric analysis of gels imaged by the densitometer was carried out using the manufacturer's proprietary software. Quantification was based on the band intensity obtained from a pure standard of the relevant protein run and imaged from each gel. Densitometric analysis of gels visualised using the Syngene G:Box ChemiHR system was completed using the ImageJ image processing software (http://imagej.net).

2.5.2 Glucose oxidase activity assay

The Amplex® Red Glucose/Glucose Oxidase Assay Kit (Thermo Fisher Scientific, UK) was used to assay for the titre of secreted recombinant glucose oxidase in culture medium, based on its enzymatic activity, according to the manufacturer's instructions. Cell cultures were centrifuged at 14,000 rpm for 2 minutes and 1 in 10 dilutions of culture supernatants in the assay buffer were prepared. Samples were assayed in triplicate in black opaque 96 well microtiter plates (Greiner Bio-One, UK) using the BioTek Synergy HT plate reader (BioTek, UK). The fluorescence from the oxidised form of the Amplex red reagent, which is produced at a rate proportional the concentration of glucose oxidase in the sample, was measured at an excitation of 530nm and detection of emitted light at 590nm. The fluorescence readings were converted to units (U) (where 1U is defined as the amount of enzyme that will oxidise 1.0µmol β-D-glucose to D-gluconolactone and H_2O_2 per minute at pH 5.1 and 30°C) using a standard curve that was derived from assaying GOx standards provided by the kit which can be found in 9.2 (Figure A1).

2.5.3 Lysozyme activity assay

The lysozyme activity assay formatted by Lee and Yang (2002) for use in microplates was used to determine the relative titre of secreted synthetic human lysozyme T70A in culture supernatants, based on its enzymatic activity. Dried *Micrococcus lysodeikticus* cells (Sigma Aldrich, UK) were suspended in 100mM potassium phosphate buffer pH 7.0 at a concentration of 0.3mg/ml prior to each assay. 50µl of culture supernatant was combined with 200µl of the *M. lysodeikticus* cell suspension, gently shaken for 1 minute and assayed with the BioTek Synergy HT plate reader (BioTek, UK). The decline in the OD_{450} of samples, caused by the degradation of peptidoglycan in *M. lysodeikticus* cell walls through lysozyme activity, was recorded over a 7 minute period, immediately following shaking, at 1 minute intervals. The lysozyme activity, calculated using the gradient of the linear plot of OD_{450} against time was used to infer the relative lysozyme titre of the sample in U/ml, where 1U is defined as the quantity of lysozyme required to reduce the absorbance by 1 mOD_{450} per minute.

2.6 Statistical methods

2.6.1 Modified Levene's test for homogeneity of coefficients of variation

The Modified Levene's test for the comparison of differences between the coefficients of variations of 2 independent groups was calculated using the SPSS statistical software package (http://www.ibm.com/analytics/us/en/technology/spss/).

2.6.2 Regression analysis

Regression analysis was performed using the Data Analysis ToolPak add-in for Microsoft Excel 2007 (www.office.com).

2.6.3 Mann-Whitney U test

The Mann-Whitney U test to compare differences in the median values between 2 groups without assuming the normal distribution of data sets was calculated using the Minitab statistical software package (www.minitab.com)

2.6.4 Student's T test

Unpaired Student's T tests without the assumption of homogeneity of variance were performed using the Data Analysis ToolPak add-in for Microsoft Excel 2007 (www.office.com).

2.6.5 ANOVA with Games-Howell post hoc test

Analysis of variance (ANOVA) coupled with the Games – Howell post hoc test for significance between multiple groups, designed for unequal/ small sample sizes that do not meet the assumption of homogeneity of variance, was performed using the SPSS statistical software package (http://www.ibm.com/analytics/us/en/technology/spss/).

3. Investigating the effect of marker selection on clonal variation in *P. pastoris*

3.1 Introduction

3.1.1 Aims and objectives

The aim of this experiment was to determine and quantify the effect, if any, that selection with zeocin has on clonal variation in recombinant protein production. Initially, unexposed cell populations and populations grown on a selective concentration of zeocin were compared to identify any significant changes in clonal variation attributable to a zeocin background.

Secondly, to determine whether high levels of clonal variation are unique to zeocin selected populations, this study directly compared zeocin selection to alternative *P. pastoris* selection methods that do not require the use of any mutagenic chemicals.

3.1.2 Selectable markers for *P. pastoris*

A fundamental requirement for the amenability of an expression system to genetic modification is an effective selectable marker to easily determine the successful insertion of target DNA into the organism, with the number of available markers for an expression organism directly reflecting its accessibility and versatility. The selective markers that have been identified and implemented in *P. pastoris* are listed in Table 6.

Table 6: List of known marker genes used for direct selection in *P. pastoris*.

Gene	Function	Standard method of selection in *P. pastoris*	Source
ADE1/ADE2	Biosynthetic pathway genes from *P. pastoris* (phosphoribosylaminoimidazole carboxylase), required for the synthesis of adenine	Auxotrophic complementation of *ade1* or *ade2* mutants – restores growth on minimal media lacking adenine	Lin Cereghino *et al.*, 2001
ARG4	A biosynthetic pathway gene from *P. pastoris*, (argininosuccinate lyase) required for the synthesis of arginine	Auxotrophic complementation of a*rg4* mutants – restores growth on minimal media lacking arginine	Lin Cereghino *et al.*, 2001
HIS4	A biosynthetic pathway gene from *P. pastoris*, (histidinol dehydrogenase) required for the synthesis of histidine	Auxotrophic complementation of *his4* mutants – restores growth on minimal media lacking histidine	Cregg *et al.*, 1985
MET2	A biosynthetic pathway gene from *S. cerevisiae* (homoserine-O-transacetylase) required for the synthesis of methionine	Auxotrophic complementation of *met2* mutants – restores growth on minimal media lacking methionine	Thor et al., 2005
URA3	A biosynthetic pathway gene from *P. pastoris*, (orotidine 5 - phosphate decarboxylase) required for the synthesis of uracil	Auxotrophic complementation of *ura3* mutants – restores growth on minimal media lacking uracil	Lin Cereghino *et al.*, 2001
bsd	Gene from *Aspergillus tererus* (blasticidin S deaminase), conferring resistance to the antibiotic blasticidin	Growth on 100µg/ml blasticidin	Kimura, Takatsuki, and Yamaguchi, 1994
hph	Gene from *Klebsiella pneumonia* conferring resistance to the antibiotic hygromycin b	Growth on 200µg/ml hygromycin b	Yang *et al.*, 2014

Gene	Function	Standard method of selection in *P. pastoris*	Source
nat	Gene from *Streptomyces noursei*, conferring resistance to the antibiotic nourseothricin	Growth on 100µg/ml nourseothricin	Nett *et al.*, 2013
sh ble	Gene from *Streptoalloteichus hindustanus*, conferring resistance to the antibiotic zeocin	Growth on 100µg/ml zeocin	Higgins *et al.* 1998
acc1	Gene from *Sorangium cellulosum*, conferring resistance to the macrocydic polyketide soraphen A	Growth on 0.02µg/ml soraphen a	Wan *et al.*, 2004
tn903kanr	Gene modified from *E. coli*, conferring resistance to the antibiotic G418	Growth on 0.5mg/ml G418	Ma *et al.*, 2009

The first marker systems to be implemented in *P. pastoris* exploited auxotrophic complementation – in which mutant strains incapable of synthesising a particular organic compound required for growth are transformed with genes encoding the enzyme required to rescue the biosynthetic pathway. The resulting transformants are able grow in media lacking the specific organic compound to be distinguishable from cells that have not incorporated the recombinant DNA. The development of the original auxotrophic complementation selectable markers in *P. pastoris*, *his4* and *arg4*, were not only inspired by molecular genetics tools in *S. cerevisiae* but enabled by them as a number of previously isolated biosynthetic gene homologues in *S. cerevisiae* retained their functionality when expressed in *P. pastoris* (Cregg et al, 1985). Currently a wide range of auxotrophic markers have been tested in *P. pastoris*, many of which have been brought to market as commercial vectors and strains, including the pPIC9/pPIC3.5 series of vectors and the *Pichia*Pink™ system from Thermofisher.

However auxotrophic complementation presents certain drawbacks, as each selectable marker can only be used in conjunction with its respective auxotrophic mutant strain, limiting compatibility between strains and vectors as well as prohibiting the use of wild type strains. As such, antibiotic resistance markers, often previously described in *S. cerevisiae*, have also been developed for *P. pastoris*. Despite their value as molecular biology tools only two dominant antibiotic resistance markers, the zeocin and blasticidin resistance genes: *sh ble* and *bsd*, were widely available and

accepted until 2008 (Kimura, Takatsuki, and Yamaguchi, 1994; Higgins et al, 1998). Previously the G418 resistance marker *tn903kan*^{r} was included in commercial vectors such as pPIC9k and pPIC3.5k, but its use was limited to secondary selection for multi-copy variants due to its low expression levels in *P. pastoris* (Scorer et al, 1994). This was remedied by Ma *et al*. (2009) through the creation of novel vectors containing *tn903kan*^{r} downstream of the strong, constitutive *P. pastoris* GAP promoter, enabling direct selection on G418. The recent addition of the resistance genes for hygromycin b and nourseothricin to its molecular tool kit further increases *P. pastoris*' utility for recombinant gene expression and opens the system up to more complex molecular biology projects requiring insertions of multiple expression cassettes (Nett et al., 2013; Yang et al., 2014).

3.1.3 Zeocin selection and *sh ble*

Zeocin is a copper-chelated, glycopeptide antibiotic belonging to the bleomycin family (Berdy, 1980), and has been successfully used as a dominant selective marker in bacteria, eukaryotic microorganisms, plants and animal cell lines. Its broad spectrum of toxicity is due to its ability to bind and introduce double stranded breaks in DNA once inside the cell. The gene product Sh ble, from *Streptoalloteichus hindustanus*, confers resistance by directly binding zeocin stoichiometrically to inhibit its function (Drocourt et al, 1990).

As one of the earliest markers described in *P. pastoris*, zeocin resistance is well established and arguably the most popular selective marker for the system. Among the reasons for its popularity is the fact that zeocin selection provides a number of advantages to expression vector design. The cross functionality of the *sh ble* gene in both eukaryotic and bacterial systems, as well as its small overall size, has permitted the development of much smaller *E. coli-P. pastoris* shuttle vectors in comparison to those utilising different selectable markers. For example the zeocin selection vectors pPICZ and pPICZα (Thermo Fisher Scientific) are approximately 3.6kb in size, in comparison to the aforementioned pPIC9k and pPIC3.5, marketed by Thermo Fisher Scientific, which range from ~9.0-9.3kb.

The 1:1 ratio in binding of a single unit of Sh ble to zeocin for its inactivation means that zeocin resistance is titratable, in which there is a direct relationship between the expression level of *sh ble* in the cell and the overall concentration of zeocin that it is capable of tolerating (Gatignol, Durand and

Tiraby, 1988). Post transformational vector amplification (PTVA) is a technique that exploits this relationship by sequentially subculturing transformants containing an integrated zeo^r expression vector into media containing increasing concentrations of zeocin, selecting clones which are more zeocin resistant, many of which have integrated multiple copies of the vector (Sunga, Tolstorukov, and Cregg, 2008). PTVA using zeocin resistance has been demonstrated to increase the probability of obtaining strains containing 10 copies or more from less than 1% through standard transformation, to 6%, with a 40% probability of obtaining 3 copy strains (Sunga, Tolstorukov, and Cregg, 2008). Since the use of multicopy strains is now a staple for increasing recombinant gene expression and for the other factors mentioned previously, zeocin selection provides a unique combination of advantages that prove its usefulness beyond its primary function as a dominant selectable marker.

Unfortunately the induction of random double stranded breaks into the host genome strongly increases the probability of introducing mutations through the non-homologous end joining pathway (NHEJ). NHEJ is a highly conserved DNA repair mechanism across a number of bacterial and eukaryotic species, including *P. pastoris*, and is directly involved in the re-ligation of double stranded breaks (Critchlow and Jackson, 1998; Daley et al, 2005; Näätsaari et al, 2012). However, since NHEJ-mediated DNA repair occurs independently of sequence homology-based repair of DNA breakages, the pathway is often error prone and additions or deletion of bases to break sites repaired by NHEJ are common (Khanna and Jackson, 2001). Whilst, in theory, the presence of Sh ble should prevent zeocin from damaging host DNA in the cell, a study by Oliva-Trastoy et. al (2005) established that an increased background of double stranded breaks continue to occur in resistant human cell lines exposed to zeocin. The evidence suggesting that Sh ble is incapable of fully inactivating its activity implicates zeocin as a potential mutagen in zeocin resistant cell lines. Should the standard zeocin selection process of Zeo^r transformants provide sufficient exposure to introduce random mutations in genes directly or indirectly involved in growth on methanol or protein production, the process would be directly responsible for artificially enhancing the divergence in the productivities between colonies transformed with zeo^r vectors and would therefore be classed as a contributor to clonal variation in *P. pastoris*.

3.2 Examining the effects of zeocin selection on clonal variation in industrial *P. pastoris* strains

To begin evaluating the contribution of zeocin selection towards clonal variation in the *Pichia* system - irrespective of the nature of the particular recombinant protein or its overall titre, three industrial strains secreting different proteins with a range of titres - summarised in Table 7, were donated by Fujifilm Diosynth (Billingham, UK) for this study.

Table 7: Summary of strains used within this section. A more comprehensive overview of strains can be found in 2.1.2.

Strain name	Parent strain	Mode of selection	Secreted protein
CLD804	NRRL-11430	Zeo^R	Aprotinin
CLD819	NRRL-11430	Zeo^R	Elafin
CLD883	NRRL-11430	Zeo^R	Trypsinogen

Using the technique described in section 2.4.2 cell lines for each strain were grown in YPD overnight and plated onto either YPD or YPD + 100µg/ml zeocin agar plates following serial dilution. The resulting colonies generated varied only in their exposure to zeocin during their growth, having otherwise originated from a homogenous cell line. A control and a zeocin exposed population of 47 colonies were randomly selected for each strain to proceed to micro expression trials. The colonies were initially inoculated and grown for ~16 hours at 30°C/ 200rpm in 5ml BMGH before resuspending in 2ml BMMH, transferring to 24 deep well plates and growing for 5 days at 28°C/ 200rpm. The final OD_{600} was measured, cultures were harvested by centrifugation and recombinant protein content was quantified by densitometric analysis of the corresponding band produced by SDS-PAGE of the culture supernatant. The ranges of specific productivities from each population are displayed as box plots (Figure 7A, 8A, 9A). The coefficient of variation (CV), calculated by dividing the population standard deviation by its mean, for specific productivity in each population was taken as a representative variable for clonal variation, and compared between control and zeocin selected populations. The coefficients of variation for each strain were plotted (Figure 7B, 8B, 9B) and the

modified Levene's test for CV was used to infer the statistical significance of any differences observed between the clonal variation in control and selected populations.

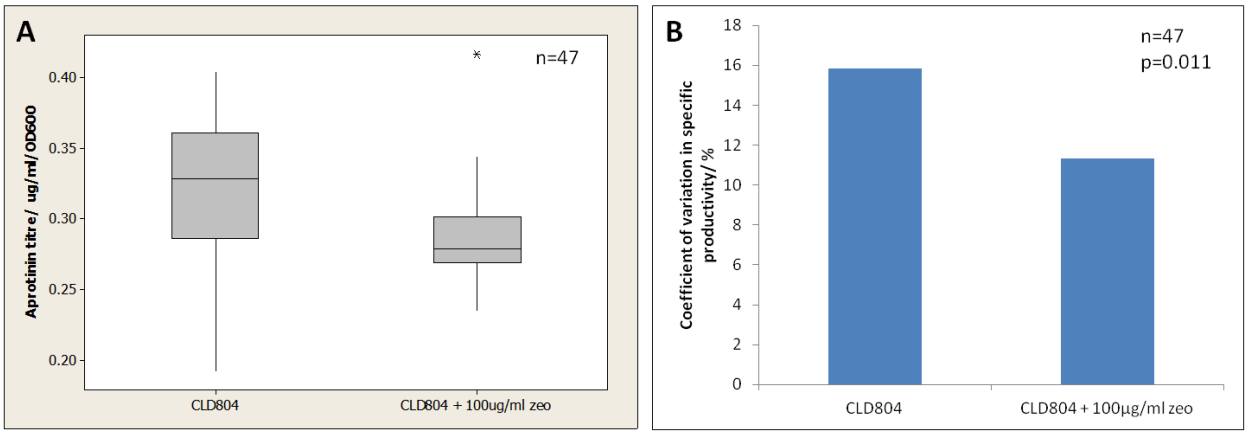


Figure 7: Graphs comparing the variation in specific productivity between clonal populations of *P. pastoris* CLD804 grown in the absence and the standard concentration of zeocin for selection, where "n" represents the number of clones in each population and "p" is the significance value generated by the modified Levene's test for the homogeneity of coefficients of variation between groups. A – Boxplot comparing range of aprotinin titres between CLD804 populations following expression over 5 days in BMM media. B – The CV for unselected and selected populations of CLD804.

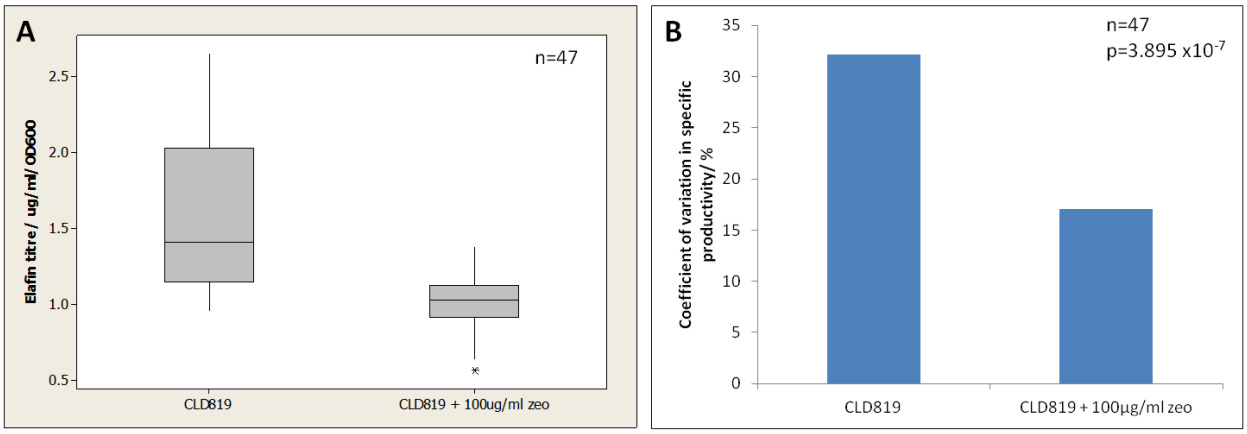


Figure 8: Graphs comparing the variation in specific productivity between clonal populations of *P. pastoris* CLD819 grown in the absence and the standard concentration of zeocin for selection, where "n" represents the number of clones in each population and "p" is the significance value generated by the modified Levene's test for the homogeneity of coefficients of variation between groups. A – Boxplot comparing range of elafin titres between CLD819 populations following expression over 5 days in BMM media. B – The CV for unselected and selected populations of CLD819.

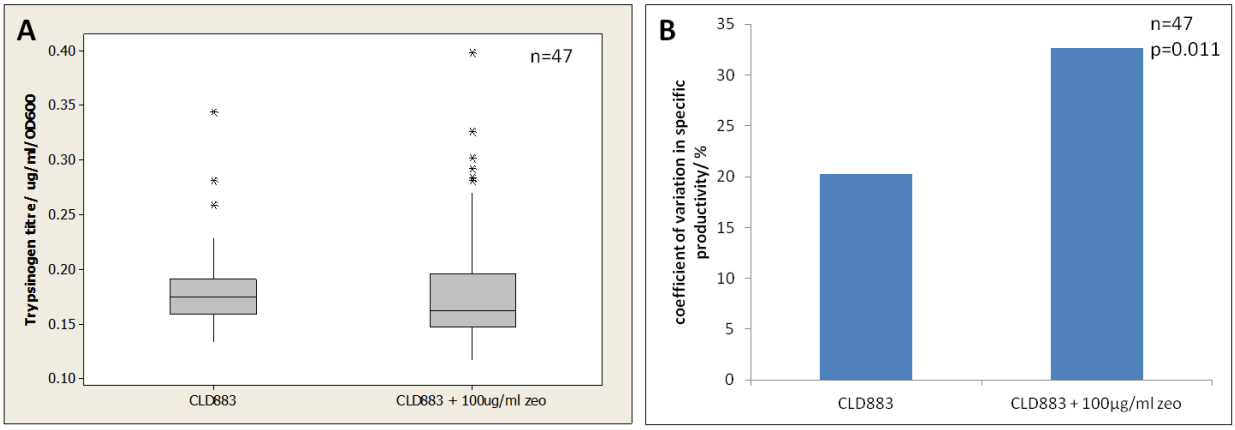


Figure 9: Graphs comparing the variation in specific productivity between clonal populations of *P. pastoris* CLD883 grown in the absence and the standard concentration of zeocin for selection, where "n" represents the number of clones in each population and "p" is the significance value generated by the modified Levene's test for the homogeneity of coefficients of variation between groups. A – Boxplot comparing range of trypsinogen titres between CLD883 populations following expression over 5 days in BMM media. B – The CV for unselected and selected populations of CLD883.

The CV observable appears to vary depending on the recombinant protein being secreted. Whilst both CLD804 and CLD883 exhibit comparable levels of variation at 15.84% and 20.4% respectively for the control populations, whilst the CV for the CLD819 control population was measured at 32.11%. The disparity in clonal variation between recombinant strains could potentially be attributed to overall recombinant protein titres, as the median specific productivity of elafin for the control population of CLD819 is approximately 4 times higher than that of CLD804 for aprotinin, and approximately 8 times higher than CLD883 for trypsinogen. Since the CV accounts for the increased degree of variation associated with larger sample averages, it therefore suggests that the higher productivity phenotype for CLD819 shows increased instability over CLD804 and CLD883. Whilst unconfirmed, as each of the strains tested within this experiment were selected by PTVA, should the titre of CLD819 exceed the other 2 strains due to it containing more copies of the elafin transgene, its increased intraclonal variation in the absence of selection could be attributed to, in part, by copy number variation between its daughter cells – the probability of which is increased in high copy number strains (Aw and Polizzi, 2013).

Contrary to the original hypothesis, for CLD804 and CLD819, less clonal variation in specific productivity was observed within the populations grown in the presence of zeocin prior to expression.

In the case of CLD804 a decreased CV for aprotinin production of 11.33% in the zeocin selected population compared to 15.84% in the control population was recorded. The difference was more pronounced for CLD819, in which the control population displayed close to double the CV of the zeocin-exposed population - 32.11% in contrast to 17.03% respectively. Both decreases were inferred to be statistically significant at the 95% confidence level. However, as can be seen from the range of titres for the populations of both species represented in the box plots (Figure 7A, 8A), the average recombinant protein titre was also lower in both of the zeocin-exposed populations. The mean specific productivity for aprotinin in CLD804 drops from 0.32µg/ml/OD_{600} in the control population to 0.29µg/ml/OD_{600} in the zeocin-exposed whereas the difference observed for mean specific productivity of elafin in CLD819 is greater – decreasing from 1.59µg/ml/OD_{600} to 1.02µg/ml/OD_{600}.

No obvious differences could be observed between the ranges of trypsinogen titres between the 2 populations of CLD883 (Figure 9A), however the CV was found to be significantly lower in the control population at the 95% confidence level, contradicting the pattern observed for CLD804 and CLD819 (Figure 9B). However, a comparison of the overall volumetric productivities for trypsinogen showed no significant differences in clonal variation between the 2 populations (data not shown) suggesting that the differences observed for the specific productivity were introduced solely by variations in final cell density.

3.3 Comparison of clonal variation between strains under zeocin and HIS4 selection

The distinct changes in variability in specific productivity observed for *P. pastoris* CLD804 and CLD819 grown in the presence of zeocin raise the question as to whether the effect is unique to selection by zeocin resistance, or if it can also be observed with alternative modes of selection. The HIS4 marker, complementating His- mutants with the native *HIS* gene to restore growth on media lacking histidine as a mode of selection, was chosen to compare against Sh ble. Selection with HIS4 requires no further addition of compounds with the potential to compromise growth or productivity, or to exhibit mutagenic properties, and so provides a suitable marker to test against *sh ble* as an alternative mode of selection.

3.3.1 Cloning

The commercial *P. pastoris* vector pPICZαB was chosen as the backbone for the vector design. pPICZαB contains a number of features required for the experimental vectors, including the *S. cerevisiae* α-mating factor pre-pro signal sequence and *sh ble*. The most notable feature of pPICZα is the presence of a number of single restriction sites flanking *sh ble*, the use of which would permit the exchange of genes into the selective marker locus, under the control of the constitutive yeast promoter - P_{TEF1} and the *CYC1* transcription terminator.

Glucose oxidase (GOx) was selected as a marker to measure specific, secreted productivity from *P. pastoris*. The ~1.8kb gene has successfully been expressed in both *S. cerevisiae* and *P. pastoris* (Malherbe, 2003) (Guo, 2010) and liquid assays for its activity are commercially available, in comparison to the proteins produced by CLD804, CLD819 and CLD883 where productivity can only be quantified by densitometry of protein electrophoresis gels. Specifically the GOx homologue produced by the filamentous fungus *Penicillium funiculosum* is reported to have a high activity and an identical optimal temperature to the growth temperature of *P. pastoris* - 30°C (Sukhacheva, 2004).

The ampicillin resistance gene (*ampR*) was selected as a bacterial selectable marker as HIS4 will not be usable in *E. coli*. A ~1.2kb region encoding the *ampR* promoter, *ampR* and a Rho-independent transcription terminator from the pUCG18 (based on pUC18) plasmid was predicted by the web-based program ARNold (Institut de Génétique et Microbiologie, Université Paris Sud), and included in the design of the vectors.

Following the above specifications 2 plasmids to compare zeocin resistance and *HIS4* complementation for their effects on clonal variation were designed and designated: pGOxZα and pGOxHα (Figure 10).

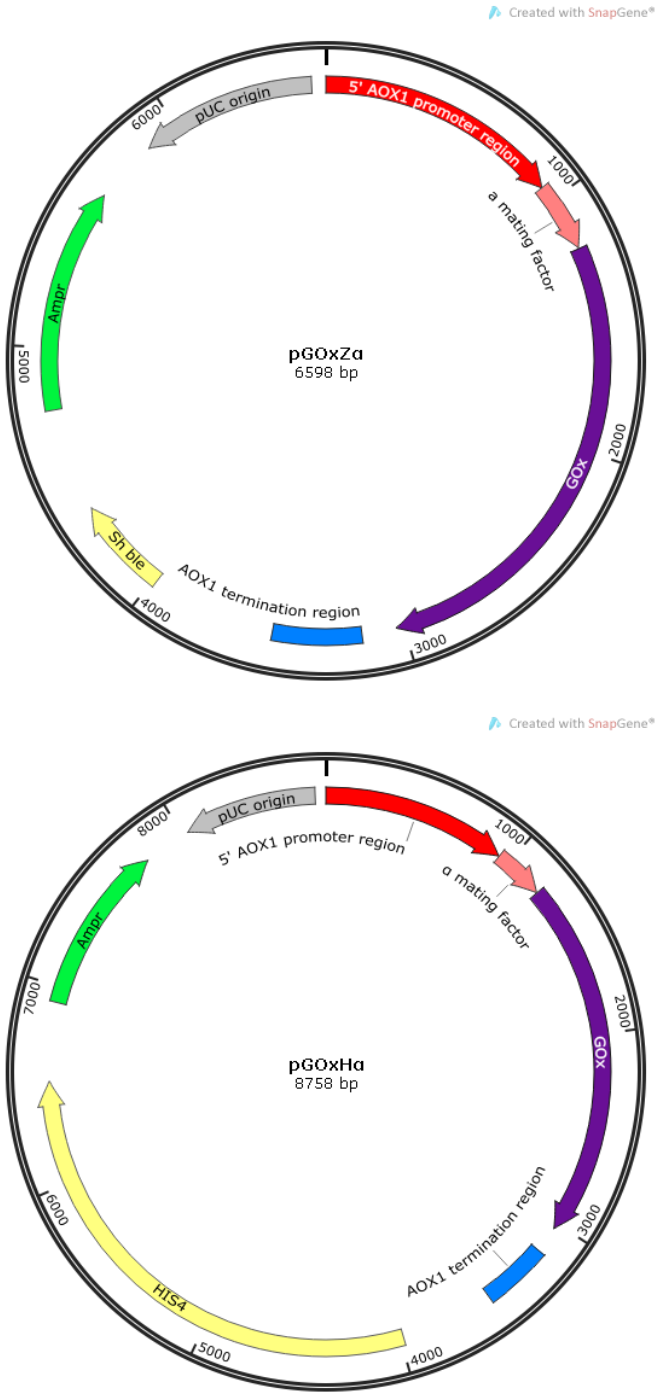


Figure 10: Vector maps of pGOxZα and pGOxHα. Aside from the selectable marker genes both consist of an identical vector backbone and expression cassette for *GOX*.

The assembly of pGOxZα was divided into the insertion of *ampR* into pPICZα to form the empty vector backbone – pAmpZα, followed by the cloning of *gox* into the multiple cloning site (MCS) to

assemble the final construct. To insert the ampR expression cassette into an independent site, pPICZα was linearized by a single digest with the restriction enzyme PciI, cutting between the pUC origin and the CYC1 terminus. The fragment encoding *ampR,* with its native promoter and terminator, was PCR amplified from pUCG18 using the Gibson primers 06-GibAmpRpαb-F and 07-GibAmpRpαb-R (9.1). The ampR expression cassette was inserted into pPICZα by Gibson assembly and cloned into *E. coli* BIOBlue. The positive transformants were identified by selection on ampicillin, colony PCR and sequencing (Figure 11).

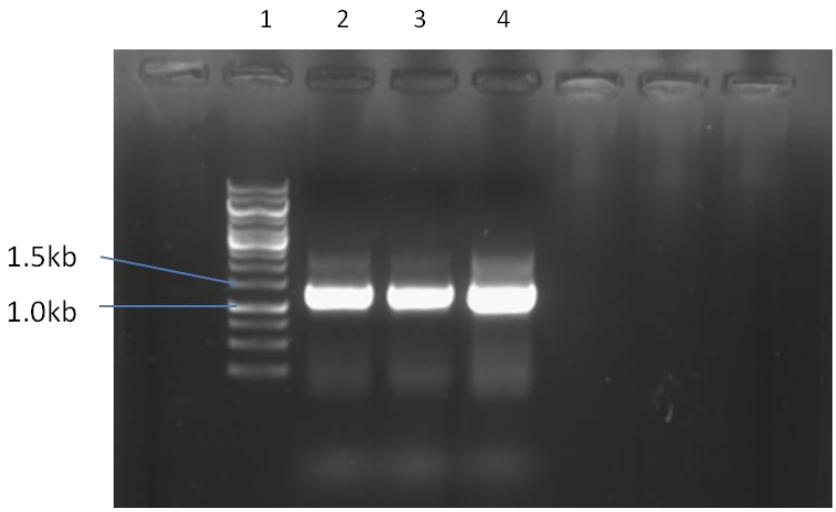


Figure 11: Colony PCR of 3 *E. coli* BioBlue colonies cloned with pAmpZα. The presence of the ~1.2kb band corresponding to the gene conferring amp^{r} indicates the successful construction of pAmpZα. Lane 1 – 1kb DNA size marker, 2-4 – PCR of individual colonies for the *ampR gene*.

GOX was PCR amplified from *P. funiculosum* genomic DNA ATCC® 11797-D2 (LGC Standards, UK), with 08-GibGOx-F and 09-GibGOx-R (9.1). Following its assembly, pAmpZα was linearised within the MCS with PstI. Gibson assembly was once again used to insert *gox* into the linearized pAmpZα, in frame of the α mating factor, cloned into *E. coli* BIOBlue, and selected on 100µg/ml ampicillin. The successful assembly of pGOxZα was confirmed by colony PCR and sequencing of the AOX1 locus (Figure 12).

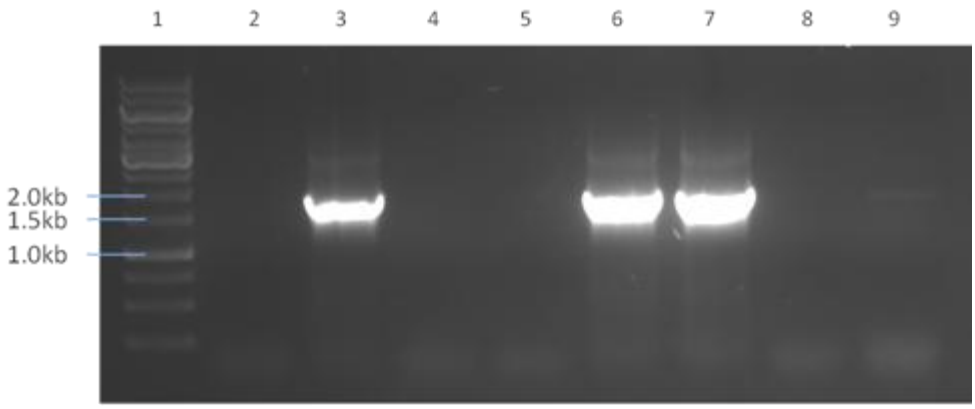


Figure 12: Colony PCR of *E. coli* BIOBlue cloned with pGOxZα. The 1.8kb band corresponding to *GOX*, found in 3 of the tested colonies indicates the successful assembly of pGOxZα. Lanes 1: 1kb DNA size marker, 2-9: PCR from 8 individual colonies.

The assembly of pGOxHα required the replacement of *sh ble* from pAmpZα with *his4*, without altering the corresponding promoter and transcription terminator flanking the gene within the vector. To remove the ~400bp *sh ble* gene from pAmpZα, a double digest was carried out with NcoI and StuI. Whereas NcoI cuts between the 5' end of *sh ble* and the *TEF1* promoter, the StuI site is located 30bp downstream of the 3' end of *sh ble*, resulting in the removal of 30bp of the *CYC1* transcription termination region during the double digest. To restore *CYC1* TT upon the insertion of *HIS4*, a reverse PCR primer for *HIS4* was designed to include the 30bp region that was excised from pAmpZα. *HIS4*, the 5' 30bp of CYC1tt and 5'/3' linker regions was successfully PCR amplified from the *P. pastoris* vector – pIB2, and Gibson assembled into the double digest of pAmpZα. The resulting *E. coli* transformants were screened by colony PCR, and sequenced with primers annealing 50 bp upstream and downstream of the predicted position of *HIS4*, to confirm its insertion as well as the restoration of the full *CYC1* terminator (Figure 13).

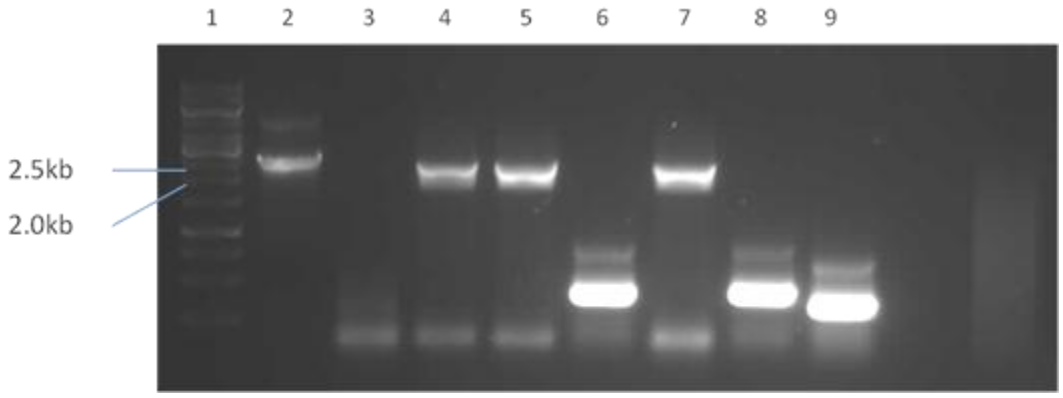


Figure 13: Colony PCR of *E. coli* BioBlue cloned with pAmpZα + *HIS4*. The presence of the ~2.5kb band (corresponding to the size of his4) indicates the successful replacement of sh ble with *HIS4* within the vector. Lanes 1 - 1kb DNA size marker, 2-9 - PCR from 8 individual colonies for *HIS4*

Following the replacement of *sh ble* in pAmpZα with *HIS4*, *GOX* was inserted into the vector MCS to complete the assembly of pGOxHα, by the same method used for the construction of pGOxZα.

To clone pGOxZα and pGOxHα into *P. pastoris* the plasmids were linearised in the *AOX1* promoter region with Sac1 and transformed into *P. pastoris* GS115 (His-) by electroporation. The resulting transformants were selected either on 100µg/ml zeocin for pGOxZα, or on MD media for pGOxHα. The new recombinant strains were named *P. pastoris* GpαGOxZ (zeoR) and GpαGOxH (His+), and are summarised for reference in Table 8.

Table 8: Summary of strains used within this section. A more comprehensive overview of strains can be found in section 2.1.2.

Strain name	Parent strain	Mode of selection	Secreted protein
GpαGOxZ	GS115	Zeo^R	Glucose oxidase
GpαGOxH	GS115	His+	Glucose oxidase

3.3.2 Expression trials of control and selected clonal populations of *P. pastoris* GpαGOxZ and GpαGOxH

Using the technique described in section 2.4.2 a single colony of GpαGOxZ was inoculated into YPD medium, grown overnight and plated onto either YPD or YPD + 100µg/ml zeocin agar plates following serial dilution. The same process was repeated for GpαGOxH except that zeocin agar plates were substituted for MD agar plates. After growth on plates a selected and control population of 24 colonies were randomly selected for each strain to proceed to micro expression trials. The colonies were initially inoculated and grown for ~16 hours at 30°C/ 200rpm in 5ml BMGY before resuspending cultures in 3ml BMMY at a normalised OD_{600} of 5.0, transferring to 24 deep well plates and growing for 48 hours at 28°C/ 200rpm. The final OD_{600} was measured, cultures were harvested by centrifugation and recombinant protein content was quantified using the Amplex Red Glucose Oxidase Assay Kit (ThermoFisher Scientific, UK) of the culture supernatant. Figure 14 displays the range and CV of specific productivities for each clonal population of GpαGOxZ and GpαGOxH.

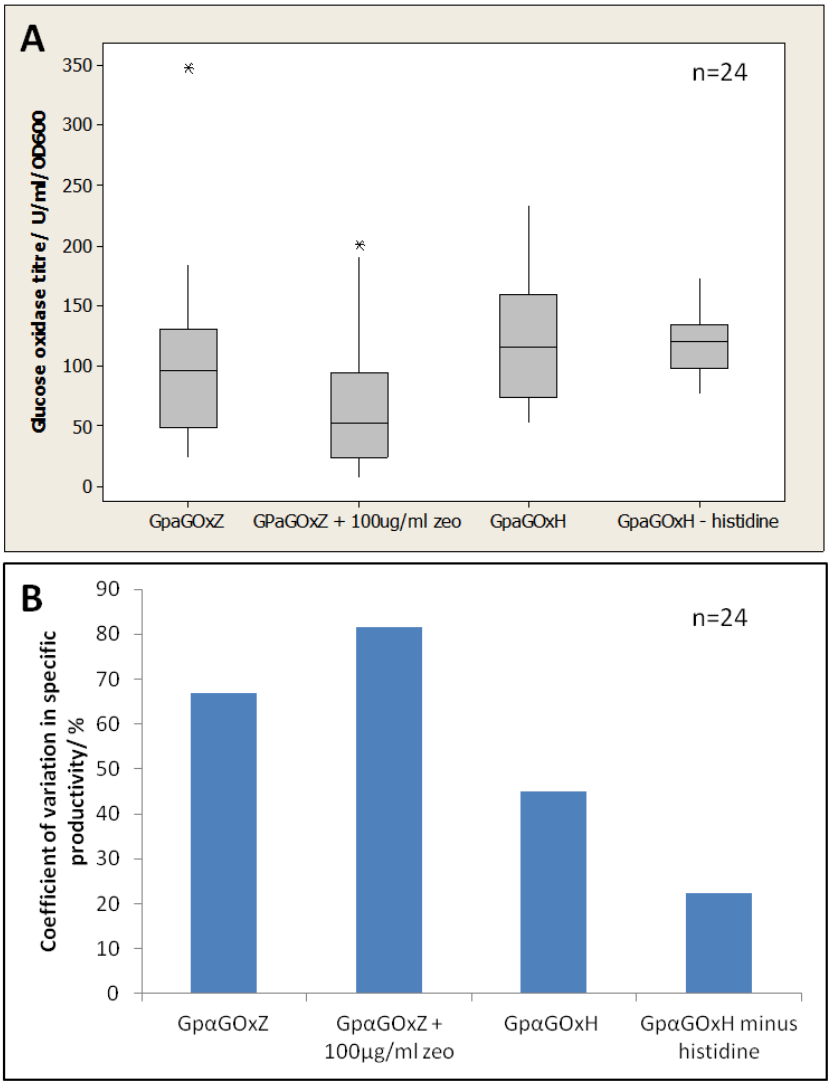


Figure 14: Graphs comparing the variation in specific productivity between clonal populations of *P. pastoris* GpαGOxZ and GpαGOxH grown in the absence and presence of the relevant selective marker, where "n" represents the number of clones in each population. A – Boxplot comparing range of glucose oxidase titres between GpαGOxZ and GpαGOxH populations following expression over 48 hours in BMMY media. B – The CV for unselected and selected populations of GpαGOxZ and GpαGOxH. The modified Levene's test was used to test for significance between coefficients of variation between unselected and selected populations of each strain, yielding p values of 0.157 for GpαGOxZ and 0.001 for GpαGOxH.

The zeocin-exposed and control populations of GpαGOxZ produced similar ranges of glucose oxidase titres, and no statistically significant differences in their coefficients of variation, suggesting that no detectable difference in clonal variation existed between them. However, under selective conditions, the GpαGOxH population possessed a CV of approximately half of the respective control population, at 22.47% compared to 44.96% for the latter. Whilst the magnitude of difference in CV is similar to

the relationship between the selected and control populations of CLD819, the mean specific productivities of both populations of GpαGOxH were roughly equivalent, in contrast to the relationship observed between selected and exposed populations of CLD804 and CLD819. This suggests that, in addition to the type of recombinant protein being produced, the effect on clonal variation could also be dependent on the mode of selection. Based on these results no clear rationale can be provided to explain how a reduction in clonal variation can occur with HIS4 selection but not under zeocin selection, other than it could provide a stronger selective constraint on factors that contribute to the specific productivity phenotype, since its gene product is an essential component in a metabolic pathway that enables both protein synthesis and cell growth in a minimal medium. Regardless, it can be concluded that the partial relationship between selection and clonal variation observed is not unique to zeocin selection but appears to apply to alternative methods of selection.

3.4 Maintenance of selection during expression does not have a significant effect on clonal variation

Having observed a significant decrease in clonal variation in specific productivity of glucose oxidase within the selected population of GpαGOxH it would be pertinent to test whether the variability within the population can be sustained, or further reduced, by maintaining selection during a longer period of the strain's use. Clonal variation places constraints on expression conditions and fermentation strategies as copy number stability and consistency in protein production begin to deteriorate as the time period of expression is increased. A possible explanation for how variability arises during extended periods of expression is that a drift in specific productivity caused by processes such as copy number loss by loop-out recombination, or clonal variation, results in a rise of new variants with productivities differing from the native strain, that continue to persist or even out-compete the original high-producing strain in fermentor culture.

Standard protocols for growth and expression of *P. pastoris* do not specify for selection to be maintained during expression trials. Yet, based on the effect observed when the initial growth of certain populations was conducted in the presence of a selectable marker, it is possible that selection during recombinant protein expression could confer stability in specific productivity over an increased period of time. If a consistent reduction in clonal variation can be achieved during expression by

maintaining a selective environment, this would implicate selectable markers as a means for improving consistency in productivity as well as increasing the stability of recombinant *P. pastoris* cell lines during fermentation.

P. pastoris GpαGOxH was chosen for the purposes of this experiment as HIS4 selection was previously demonstrated to produce a clonal population with significantly reduced variability in productivity. Selection by HIS4 complementation would also be more viable and simpler to maintain as a selective condition over the course of an expression trial compared to zeocin resistance. Specifically, the HIS4-selected population from 3.3.2 was selected to test the hypothesis that applying marker selection during its expression trial would either maintain the population's minimal degree of variability or decrease it further.

Each of the 24 cell stocks comprising the selected GpαGOxH population were inoculated in duplicate into both 5ml BMGY and BMG (- histidine) and grown for ~16 hours at 200rpm, 30°C. Cultures growing in BMGY were then passaged into BMMY, and those grown in BMG were passaged into BMM to induce expression of glucose oxidase, all at a starting OD_{600} of 5.0. Expression in 3ml culture volumes was continued for 48 hours at 28°C, shaken at 200rpm in 24 deep well plates and the variation in glucose oxidase titre was quantified as described in 3.3.2 (Figure 15).

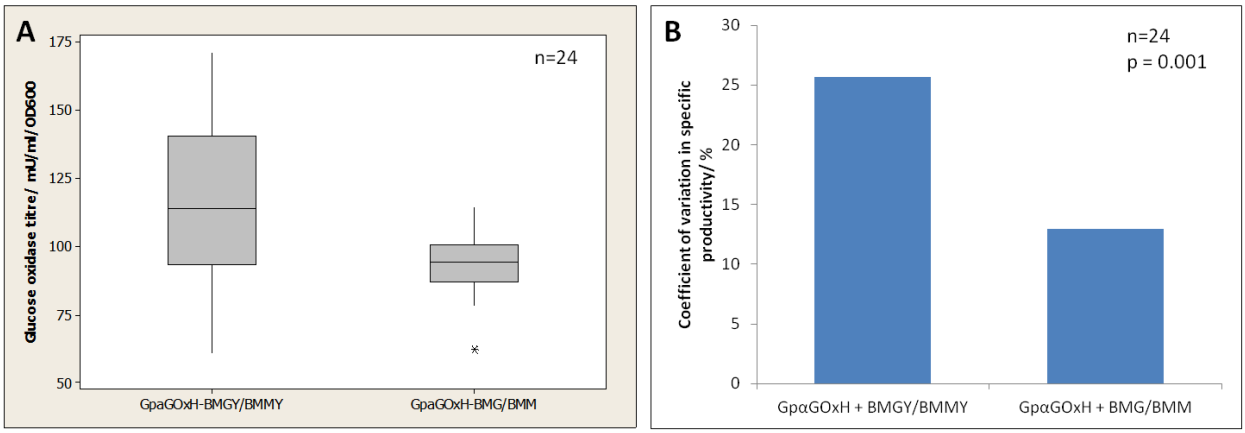


Figure 15: Graphs comparing the variation in specific productivity between a clonal population of *P. pastoris* GpαGOxH following a 48 hour expression trial in BMMY or BMM media, where "n" represents the number of clones in each population and "p" is the significance value generated by the modified Levene's test for the homogeneity of coefficients of variation between groups. A – Boxplot comparing range of glucose oxidase titres between expression on the standard expression media BMGY/BMMY and selective media BMG/BMM. B – The corresponding CV for specific productivity under each condition.

Selective conditions were imposed on the cultures grown and expressed in the minimal versions of the media used, i.e. BMG and BMM, as both lack sources of histidine. The results from Figure 15 show that, when grown and expressed in minimal media lacking histidine, variability in glucose oxidase specific productivity is significantly lower compared to when the same population is grown in non-selective rich media. The value obtained for the CV of GpαGOxH grown in BMGY/ BMMY is approximately equivalent to the previous CV obtained for GpαGOxH under identical conditions in 3.3.2: 25.7% and 22.47% respectively. However when expression was repeated in BMG/BMM the CV decreased significantly to 12.97%.

However, the large disparity between the final cell density of cultures grown in rich media and minimal media, and its subsequent effect on the overall volumetric productivity introduces a number of additional variables that could affect the validity of comparing specific productivities in BMMY and BMM (Figure 16).

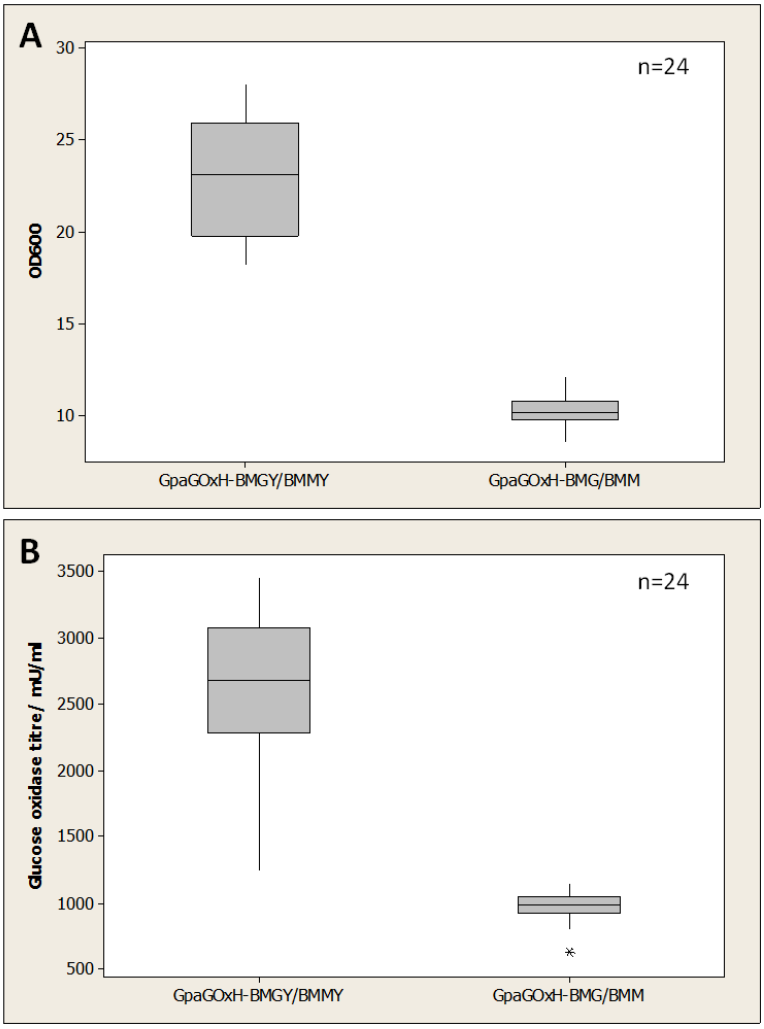


Figure 16: Boxplots comparing the final cell density (A) and volumetric productivity of glucose oxidase (B) between expression on BMGY/BMMY and selective media BMG/BMM.

Therefore the experiment was repeated but with BMGY and BMMY substituted by BMG and BMM supplemented with histidine (BMGH/BMMH) as the non-selective condition (Figure 17).

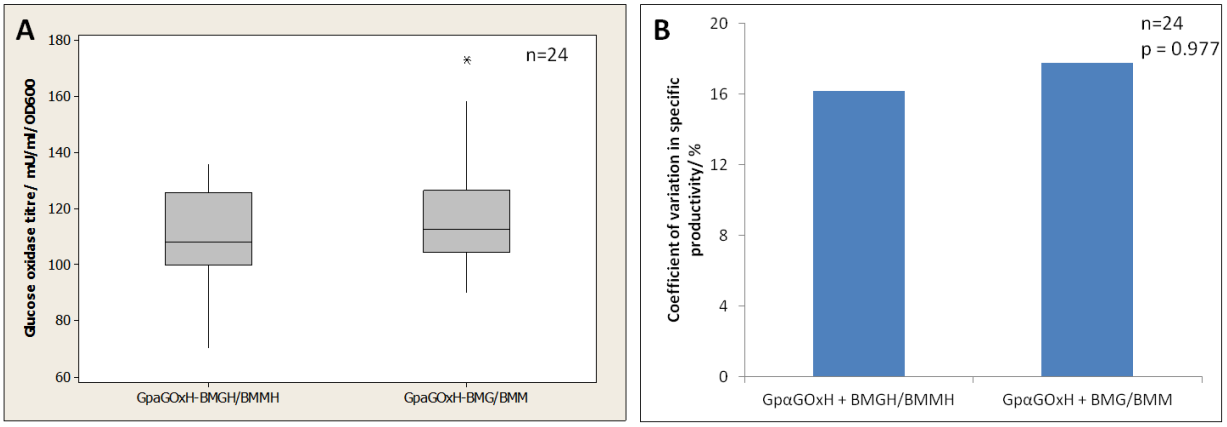


Figure 17: Graphs comparing the variation in specific productivity between a clonal population of *P. pastoris* GpαGOxH following a 48 hour expression trial in BMMH or BMM media, where "n" represents the number of clones in each population and "p" is the significance value generated by the modified Levene's test for the homogeneity of coefficients of variation between groups. A – Boxplot comparing range of glucose oxidase titres between expression on the standard minimal expression media BMGH/BMMH and selective media BMG/BMM. B – The corresponding CV for specific productivity under each condition.

Due to its similarity in composition the growth rate of *P. pastoris* in BMGH/BMMH was similar to its growth in BMG/BMM and so the final, average titres of glucose oxidase achieved in both media types are more comparable. When BMGH/BMMH was used as a non-selective condition, there was no significant difference in CV to GpαGOxH expressed under selective conditions. The significant reduction in CV observed for clonal populations of GpαGOxH grown in BMG/BMM compared to BMGY/BMMY therefore suggests that the causal factor for the difference observed is unrelated to marker selection. Regardless, the increase in variability within the GpαGOxH population grown and expressed in rich media cannot be completely explained as an inherent increase due to the larger overall productivity, as the CV accounts for the sample mean, allowing for the standardisation of data sets with widely differing mean values. This suggests that clonal variation increases in response to increases in cell growth and volumetric productivity in culture in a non-linear fashion, and is significantly minimised when growth is limited during expression.

3.5 Comparison of gene copy number variation between selected and unselected clonal populations

One of the major contributors towards variation in recombinant protein productivity between the daughter cells of a recombinant strain is the deviation in the number of copies of the target gene in the host genome, brought about by homologous recombination events within the locus of integration. To determine whether the differences in clonal variation between the clonal populations secreting glucose oxidase used in this study can be attributed to variation in *GOX* gene copy number, the relative copy number for each clone was measured.

The clonal populations used in the study included both the selected and non-selected populations for GpαGOxZ and GpαGOxH, derived in 3.3.2. The 24 clones comprising each population were first grown individually in YPD, YPD + 100µg/ml zeocin or MD, depending on the original growth conditions of the population. Genomic DNA was then purified from each colony as described in section 2.3.15 and copy number was measured by qPCR using primers specific to *gox* (40-qGOx2-F, 41-qGOx2-R) using beta-actin (*act1*) as a reference for a gene located in the genome as a single copy. The range of *GOX* copy numbers for each population is illustrated in Figure 18.

The distribution of *GOX* copy numbers in both the selected and control populations of GpαGOxH appears to be similar, with the majority of clones appearing to contain a single copy. This eliminates copy number variation as a potential explanation for the reduction in clonal variation in GpαGOxH when grown under selective conditions, implicating a variation in the promoter strength of *AOX1* between clones or in processes indirectly affecting *GOX* expression in response to marker selection.

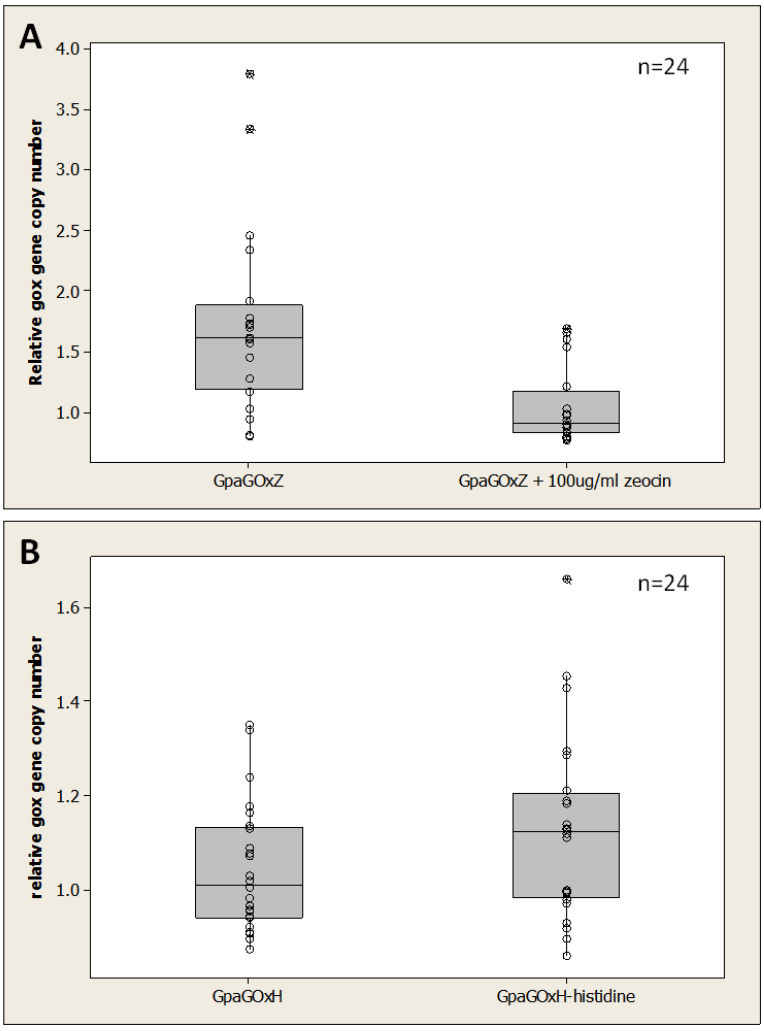


Figure 18: Boxplots showing the range of copy numbers of the *GOX* gene, measured by qPCR of previously tested clonal populations and normalised to the amplification of the act1 gene, with points representing the copy number of individual clones within each population overlaid. "n" is the number of clones plotted for each population. A – Comparison in copy number range of gox between clonal populations of *P. pastoris* GpαGOxZ raised in the absence of selection or 100µg/ml zeocin. B – Comparison in copy number range of gox between clonal populations of *P. pastoris* GpαGOxH raised in the absence of selection or media lacking histidine.

A considerable difference was observed between the zeocin-selected and control populations of GpαGOxZ, with a wider distribution, skewed further towards copy numbers higher than 1 in the control population. Conversely the population raised under zeocin selection exhibited a similar range of copy numbers to the GpαGOxH, consisting predominantly of single copy clones and 4 clones potentially containing 2 copies of *GOX*. The occurrence of higher numbers of multi-copy clones in a population raised in the absence of selection is unexpected, primarily because a selective background

favouring gene copy multimerisation is not being provided under any of the conditions tested. Despite the absence of a selective pressure favouring multiplication of the vector a clear drift in copy number has occurred in non-selected GpαGOxZ. Secondly, in spite of the disparity in the range and average copy number of clones within the 2 GpαGOxZ populations, it does not appear to affect GOx productivity as no significant difference in clonal variation of GOx titre was observed between them. The degree by which increasing the target gene copy number in *P. pastoris* affects recombinant protein production is context dependent and hinges on properties specific to the recombinant protein in question (Zhu et al, 2009B; Zhu et al, 2011; Aw and Polizzi, 2013). It is probable that an increase in copy number between the minimum of 1 and the maximum of 4, found in the control GpαGOxZ population, does not correlate with a strong increase in GOx production or that GOx production varies highly between clones with equivalent copy numbers such that a high producing single copy clone compares favourably with a 4 copy clone. Whilst the observed difference in *GOX* copy numbers between the control and selected populations of GpαGOxZ does not appear to affect clonal variation in productivity it could potentially explain the differences in clonal variation found in the zeocin-selected populations of CLD804 and CLD819.

3.6 Discussion

Previous studies exploring clonal variation in *P. pastoris* have not established a standard method for quantifying and comparing variability in protein productivity, presenting an initial challenge to analysing clonal variation empirically. The coefficient of variation, used commonly as a measure of variability of data within distributions, provides possibly the most representative measure for clonal variation. Given a set of data points for any given sample the coefficient of variation is simple to calculate and, more importantly, it is possible to test for the homogeneity of coefficients of variation belonging to 2 or more groups through a modification of the Levene's test: an inferential statistic originally designed to test the equality of variances. The Levene's test is particularly robust for inferring statistical significance over other more commonly used tests for variance, such as the F test, as it does not make the assumption that the data fits a normal distribution - which is relevant since the distribution of protein titres from *P. pastoris* is unknown (Levene, 1960). However, unlike variance, the coefficient of variation is dimensionless and can therefore provide a standardised measure for dispersion for highly diverse data sets. Valid comparisons between distributions varying

strongly in their means, measured variables and magnitude are possible with the coefficient of variation and so it is ideally suited for measuring variation between clonal populations with differing average titres or even the titres of different recombinant proteins if required.

Initial research investigating the effect of zeocin selection on clonal variation in *P. pastoris* revealed that variation in specific productivity for 2 of 3 recombinant proteins tested was reduced when clonal populations were grown in the standard selective concentration of zeocin. Whilst this directly contradicted the original hypothesis that the use of zeocin was contributing to clonal variation the fact that no significant difference in clonal variation could be observed for zeocin – selected clonal populations secreting trypsinogen or glucose oxidase suggests that the effect is not consistently replicable. For the purpose of discovering if the reduction in clonal variation of populations grown under selection was a unique feature of zeocin selection the recombinant strains GpαGOxZ and GpαGOxH, sharing an identical parent strain and both secreting recombinant glucose oxidase but differing in their selectable markers, was created. Whilst the mechanism of selection for HIS4 in GpαGOxH functions in an entirely separate fashion from zeocin resistance, a clonal population of GpαGOxH under selection exhibited a reduction in clonal variation of a similar scale to the previously tested strains. Hence it is highly likely that raising a population with less clonal variation is achievable through multiple selectable markers, and is not limited to zeocin resistance. Nevertheless, the application of HIS4 selection throughout expression is unsuccessful at maintaining a significantly lower variability in titre of GOx compared to using non-selective media for expression of the same clonal population. This suggests that selection only exerts an effect on the diversity of protein secretion during the formative stages of the population. For the purposes of this study clonal populations were formed selecting individual colonies, forming allegedly from single cells, on agar plates. One possible explanation is that selection promotes the proliferation of individual cells with certain phenotypes pertaining to growth or expression of the selectable marker, while cells that are compromised or grow poorly in response to selection do not form colonies. Without any hypothetical penalties imposed by selection a greater extent of drift in growth and expression phenotypes could be able to occur, resulting in a population of colony forming units with a greater diversity in growth rate and gene expression.

Surprisingly the range and average in copy number of *GOX* per clone was higher among a population of GpαGOxZ raised without zeocin selection, since there was no selective pressure present to stimulate an increase in copy number. Conversely a larger proportion of clones from the population

grown with zeocin selection contained only a single copy of *GOX*. Whilst PTVA utilises selectable markers to increase the probability of obtaining clones with high target gene copy numbers possibly only the more excessive concentrations of zeocin used in this practice, ranging up to a maximum of ~3.0mg/ml (Sunga, Tolstorukov and Cregg, 2008) in comparison to the standard selective concentration of 100µg/ml applied in this study, provides a sufficient selective advantage for multi-copy clones. During exposure to low concentrations of antibiotic that do not require an increased gene dosage to confer resistance, the overexpression of the respective selectable marker could theoretically provide a selective disadvantage to high copy number strains in culture. In this case, the over-production of redundant quantities of Sh ble in high copy strains would place a stronger burden on protein biosynthetic pathways and impeding the translation of other native proteins as the more abundant *sh ble* mRNA would increasingly occupy the limited number of ribosome binding sites within the cell. This does not explain how the occurrence of multi-copy clones in an un-selected population is higher, although the phenomenon itself could provide an explanation as to how the average recombinant protein titres were higher in both non-selected populations of *P. pastoris* CLD804 and CLD819.

Regardless of the unexpected, complex relationship observed between marker selection and clonal variation, the key conclusion that can be taken from the study is that there is no evidence to support that the use of zeocin as a selectable marker contributes significantly towards clonal variation. Selection by zeocin resistance can therefore be discounted as a candidate for inducing clonal variation within the *Pichia* system.

4. Correlating variation in growth and the expression of native genes linked to specific productivity with clonal variation

4.1 Introduction

4.1.1 Aims and objectives

This study quantified the recombinant protein titre from a large subset of clones taken from a single copy cell line in order to identify a basal, native degree of clonal variation representative of the *P. pastoris* system.

Variation in cell growth and the transcription of candidate genes with hypothetical links to recombinant protein production were quantified, and regression analysis was used to determine their impact, if any, on clonal variation within the sample.

4.1.2 Examining clonal variation in factors beyond protein production

Clonal variation in *P. pastoris* is responsible for complications in upstream processing as well as drift in productivity during extended fermentation trials, at least in multicopy strains (Higgins and Cregg, 1998; Zhu et al, 2009B; Love et al, 2010). As such there is an interest in finding and removing any causal factors that exacerbate variability between clones. However, with the exception of copy number, the absence of factors with obvious effects on clonal variation within standard *P. pastoris* expression protocols suggests that clonal variation could occur, at least in part, due to native sources of variability between cells.

Compared to *S. cerevisiae P. pastoris* appears to be significantly recombinogenic (Cregg and Russell, 2007) and it is this property which has been associated with the poor stability of multicopy clones (Mansur et al, 2005; Aw and Polizzi, 2013). Since the majority of multicopy clones in *P. pastoris* are generated by integrating tandem repeats of an expression vector within a single locus, the probability of recombination events occurring within the specified locus and disrupting the recombinant

sequence is increased. However homologous recombination itself is an agent for inducing mutations and driving genetic variation within eukaryotes through a number of mechanisms - including cross-over events during meiosis and genome rearrangements (Cooper, 2000). Should the degree of variability observed within the locus of integration in *P. pastoris* be repeated in other loci within the genome it would be rational to predict a degree of natural diversity arising within a population for a number of traits that could affect individual protein productivity. Whilst information on natural variation in *P. pastoris* is limited, a study from Love et al (2010) was not only capable of resolving significant variation in the rate and overall secretion of a human Fc antibody fragment between isolated, single cells, but that it could be correlated in part with more than a 2-fold range in cell doubling time between tested sub-populations. The results indicate that a divergence in recombinant protein secretion and growth rate, a complex phenotype with ramifications for protein production, is occurring at the single cell level.

Typically clonal variation in *P. pastoris* is only characterised by the range of final protein titres obtained from clones within a tested sample. However, extending the scope of clonal variation to by examining variability in relevant native factors such as growth rate, gene transcription or ER stress response elements and correlating the results with protein productivity will provide further information on how variability in protein productivity arises in *P. pastoris* in addition to identifying phenotypes that predispose clones to produce optimal protein titres. Furthermore if any strong correlations with protein titre can be identified, the tested native factors could inform the development of markers to screen for high producing clones without the need for expression trials.

4.2 Establishing a distribution of recombinant protein titres from a single copy clonal population

Having previously set up a simple and accurate protocol for measuring recombinant glucose oxidase secretion (chapter 3.3.2) *P. pastoris* GpαGOxZ was selected as the starting strain from which to conduct the study. In order to minimise the risk of obtaining a clonal population with a significant divergence in gene copy numbers, distinguishing between native clonal variation and variation due to heterogeneity in copy numbers, a GpαGOxZ clone confirmed to contain a single copy of *gox* in chapter 3.5 was used as the parent strain. A sample of 40 discrete colonies was randomly selected

from a series of clones grown on YPD + 100µ/ml zeocin plates following the procedure described in 2.4.2 to proceed with micro expression trials.

A 48 hour expression trial in 3ml BMMY in 24 deep well plates was conducted, after which the final titre of GOx was measured, and total RNA was isolated for each clone. The distribution of specific productivities within the sample is illustrated in Figure 19.

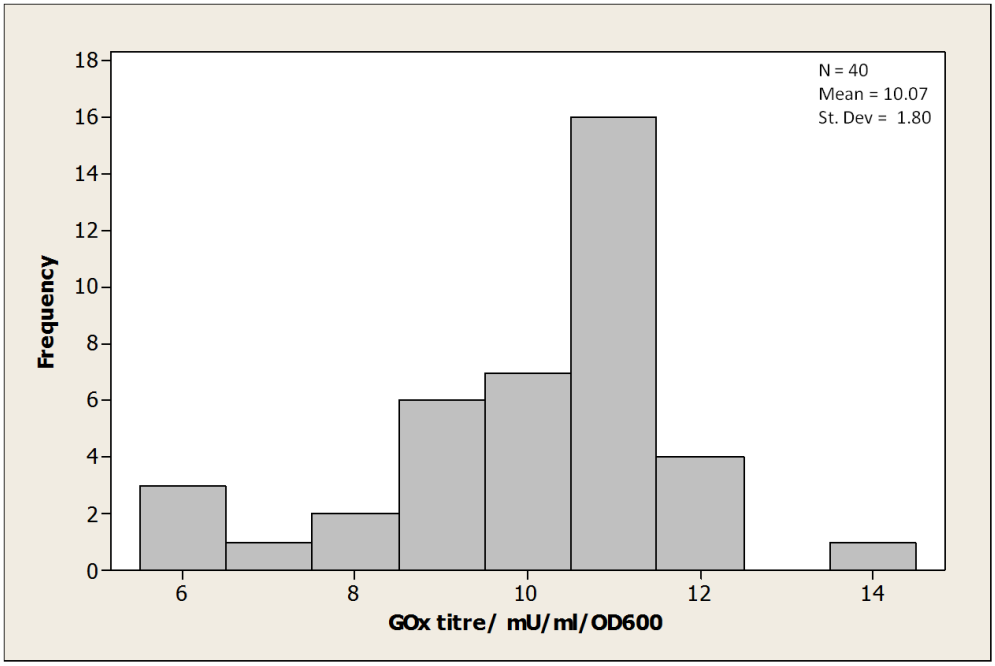


Figure 19: Distribution of the specific productivities of GOx within a clonal population derived from a single copy variant of GpαGOxZ, following a 48 hour expression trial. The mean specific productivity, standard deviation and sample size (N) are displayed.

The Anderson-Darling test for normality indicates that the range of specific productivities obtained for the sample of 40 clones does not belong to a normal distribution at a 99.5% confidence level. The majority of clones exhibit a narrow range of productivities, with greater than 50% of the sample producing GOx titres between 9.5 and 11.5 mU/ml/OD_{600}. In spite of this, significant clonal variation within the sample group is apparent when examining the lowest and highest producers within the sample, in which a greater than a 2-fold difference in GOx titre even though both subsets of clones originated from a homogenous, single copy cell line. As such, the nature in which the clonal population was raised, and the magnitude of the difference in productivity does suggest the

occurrence of stochastic events influencing recombinant protein secretion in independent clones that cannot be accounted for by simple biological variation.

4.3 Analysing the effect of total growth and growth rate on GOx productivity

Contradictory evidence exists for the effect of growth rate in culture on overall recombinant protein titre in *P. pastoris*. A transcriptomic analysis of *P. pastoris* identified the upregulation of numerous genes involved in ribosome biogenesis, protein synthesis and translocation when grown under conditions promoting higher specific growth rates, implicating a high growth rate as beneficial towards protein production (Rebnegger et al. 2014). Conversely, in the aforementioned single cell analysis of *P. pastoris* secreting constitutively expressed Fc fragment, a direct trend between specific productivity and cell doubling time was observed, suggesting that overall productivity is, in fact, inversely proportional to the growth rate of independent cells in culture (Love, 2010). However the same study also noted a subgroup, characterised by a rapid average growth rate, producing higher titres of Fc in comparison to other tested subgroups. Whilst the suitability of faster growing clones to recombinant protein production remains unclear its association with specific productivity of cells in culture creates a case for growth rate during expression as a predictor variable for correlation studies.

Both specific and volumetric productivities were initially correlated with the final OD_{600} recorded following the 48 hour expression of the 40 clones tested within the sample (Figure 20).

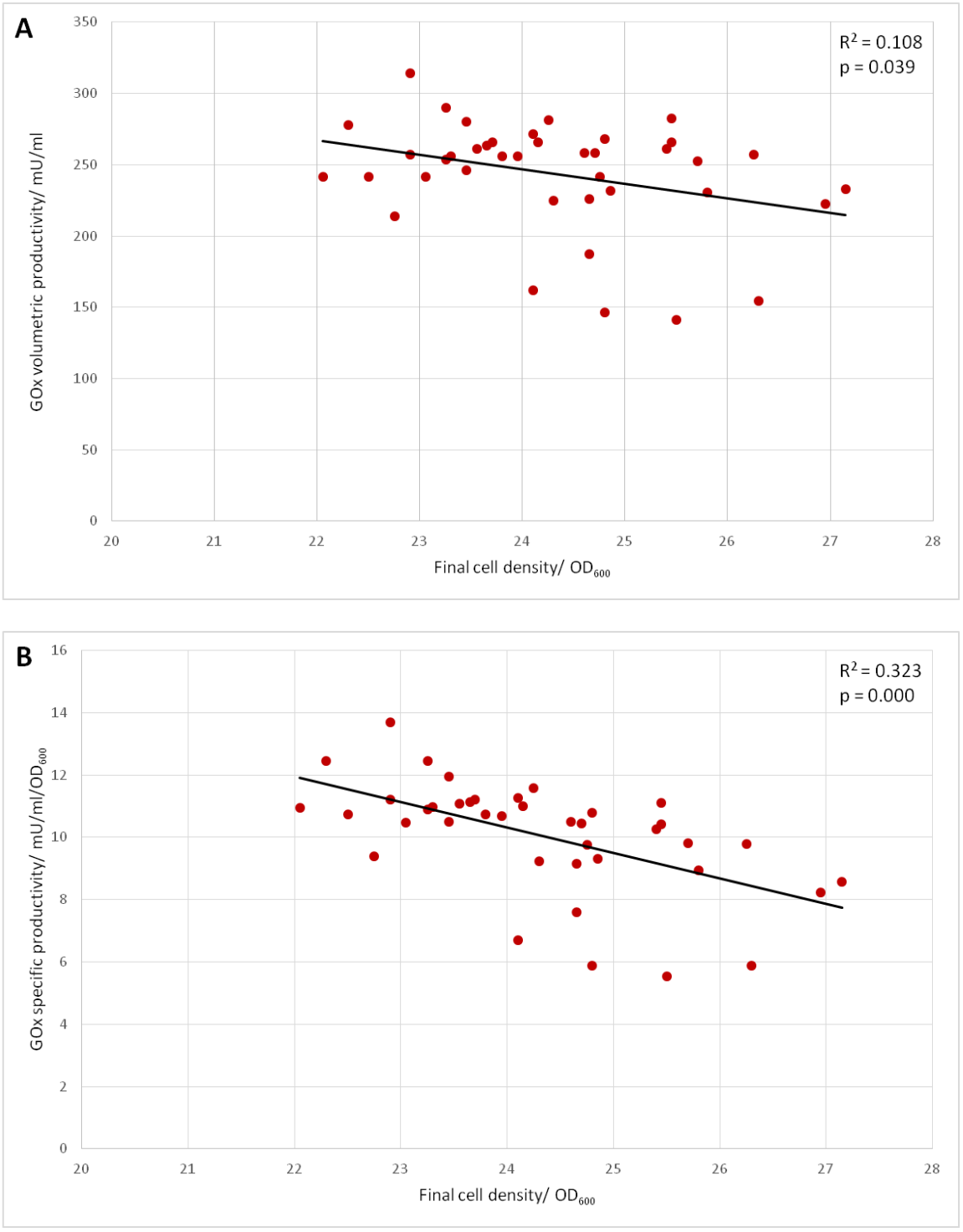


Figure 20: Linear regression between the volumetric productivity (A) or specific productivity (B) of GOx and the final cell density of cultures following a 48 hour expression trial. The R^2 value and significance value (p) from the regression analysis between the variables are displayed.

Simple linear regression was conducted in each case to determine the strength of the correlation, and coupled with regression analysis to evaluate their statistical significance. The results of the regression analysis confirmed that the weak, negative correlation found between the final cell density of clones in culture and specific productivity was significant ($p<0.05$). However caution should be taken when interpreting the relationship between final cell density and specific productivity, as the former is factored into the formula for calculating specific productivity, creating a potential bias in the strength of the correlation. Therefore, in this case, examining the correlation with volumetric productivity would provide a more accurate representation of the effect of final cell density on clonal variation. As expected the strength of the correlation between final cell density and volumetric productivity is weaker and has a lower statistical power in comparison to the equivalent correlation for specific productivity. However the reduced R^2 value remained statistically significant within a 95% confidence limit ($p<0.05$).

The OD_{600} recorded upon culture harvest however does not provide a wholly representative value to describe the growth phenotype as, after a 48 hour period, the measurement would have most likely been taken during the strain's stationary phase of growth. The growth rate during exponential phase, usually calculated through growth curve analysis, is more commonly used to denote the growth phenotype and viability of a strain in culture, and is impossible to derive from the final OD_{600}. To identify any differences in growth rate between highest and lowest producing clones from the tested sample, and its connection to productivity, the sample was binned into quartiles based on the specific productivity for each clone, with the 1^{st} (Q1) and 4^{th} (Q4) quartiles, being selected for growth curve analysis (Figure 21).

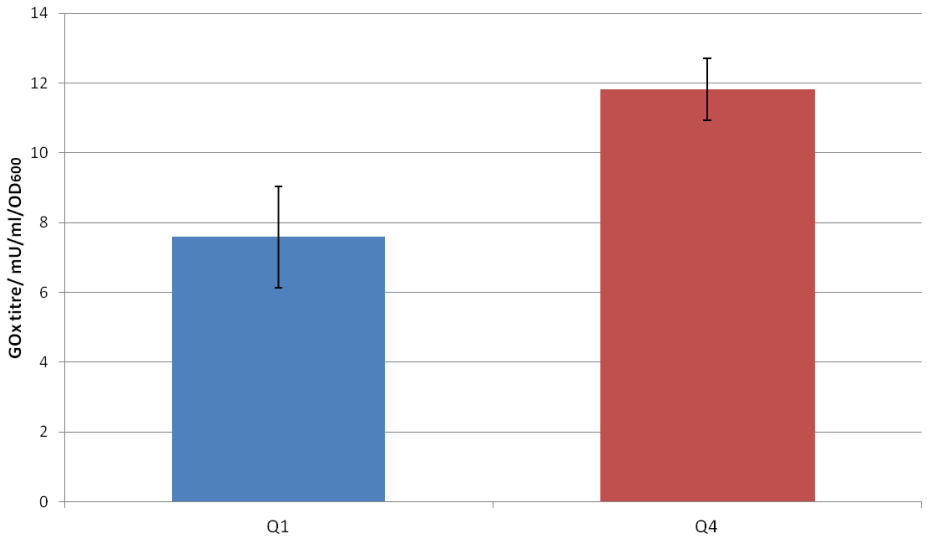


Figure 21: Average titre of GOx for sampled clones within the 1st and 4th quartiles for specific productivity. Error bars represent the standard deviation of 10 clones making up each quartile.

Growth curve analyses were conducted for Q1 and Q4, in which each clone was grown for ~16 hours in 5ml BMGY before resuspending to a final OD_{600} of 0.1 in 25ml BMMY in 250ml baffled flasks. The cultures were grown at 30°C, and OD_{600} was measured at 2 hour intervals throughout exponential growth until the beginning of stationary phase. The recorded OD_{600} was plotted for each clone, and the points encompassing the linear portion of exponential growth were used to calculate the doubling time – equivalent to the growth rate of each clone in culture during methanol-induced expression. The doubling time was calculated using the formula:

$$Doubling\ time = \frac{t * \log(2)}{\log(Final\ cell\ density) - \log(Initial\ cell\ density)}$$

In which the initial cell density represents the OD_{600} recorded at the start of linear growth, final cell density is the OD_{600} recorded upon the end of linear growth and t is the time period (in hours) between the 2 points.

To account for any differences in conditions from the initial expression trial, the final OD_{600} and glucose oxidase titres for each culture were measured again once stationary phase had been entered.

The new values for volumetric and specific productivity were then plotted against the doubling times calculated for each culture (Figure 22).

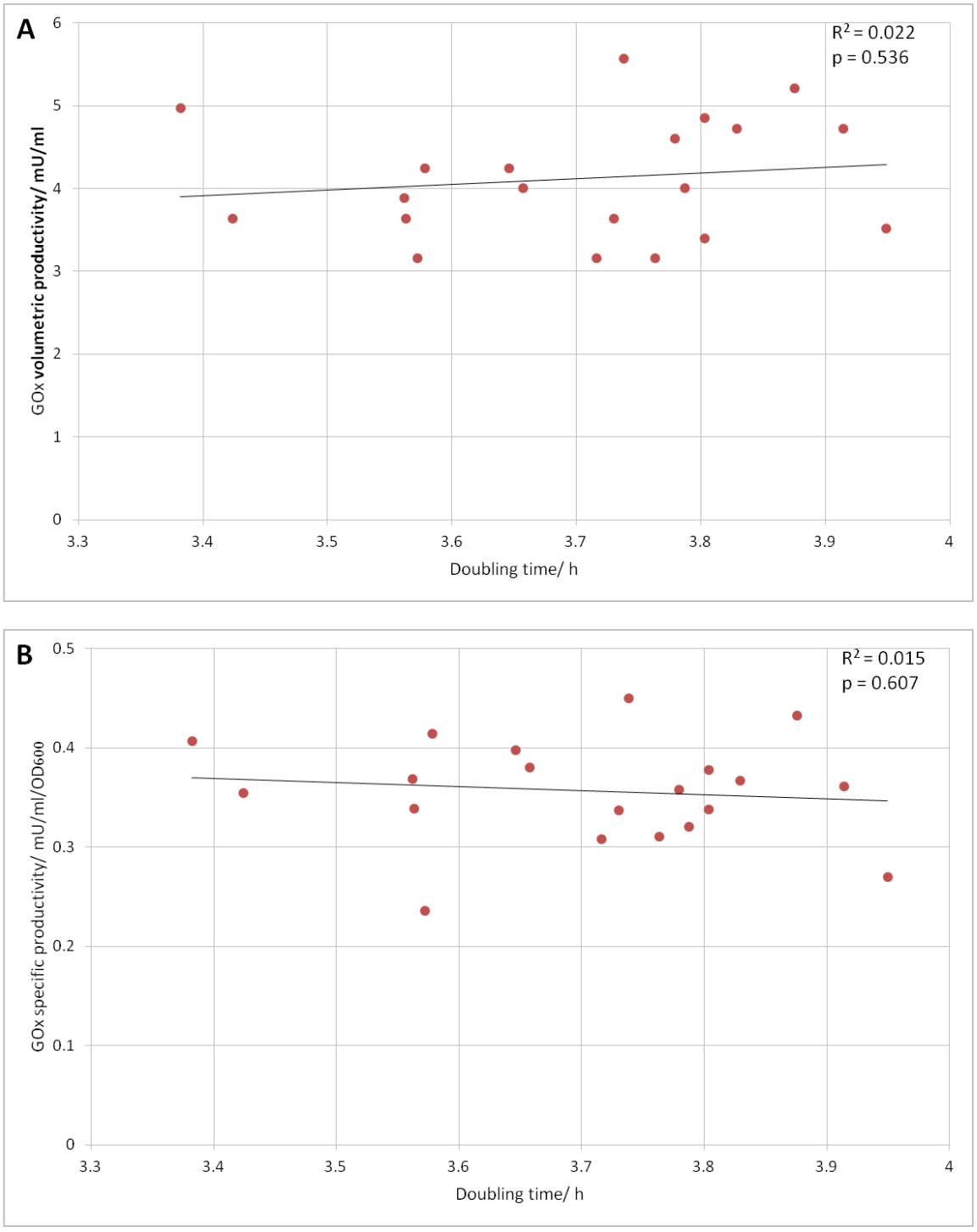


Figure 22: Linear regression between the volumetric productivity (A) or specific productivity (B) of GOx and the doubling time of clones within Q1 and Q4 for specific productivity of the original test sample. The R^2 values and significance values (p) from the regression analysis between the variables are displayed.

No significant relationship between either volumetric or specific productivity, and doubling time could be observed within the limits of the test. However, due to the lower starting cell density and time period required for growth curve analysis in comparison to a 48 hour expression trial, the GOx titres recorded for clones within Q1 and Q4 were severely low to the point in which the activities recorded by the GOx liquid assay were measured at an extremely low point on the standard curve, indicating the possibility that insufficient expression to differentiate higher producers from lower producers based on the assay alone has occurred. The enzyme activities recorded could therefore be unrepresentative of clone productivity under the conditions imposed during the original expression trial, with the added risk of the differences in titres being obscured by the background noise from the assay at such reduced activities. Bearing this observation in mind the doubling times for each clone were grouped based on the original quartiles and compared (Figure 23).

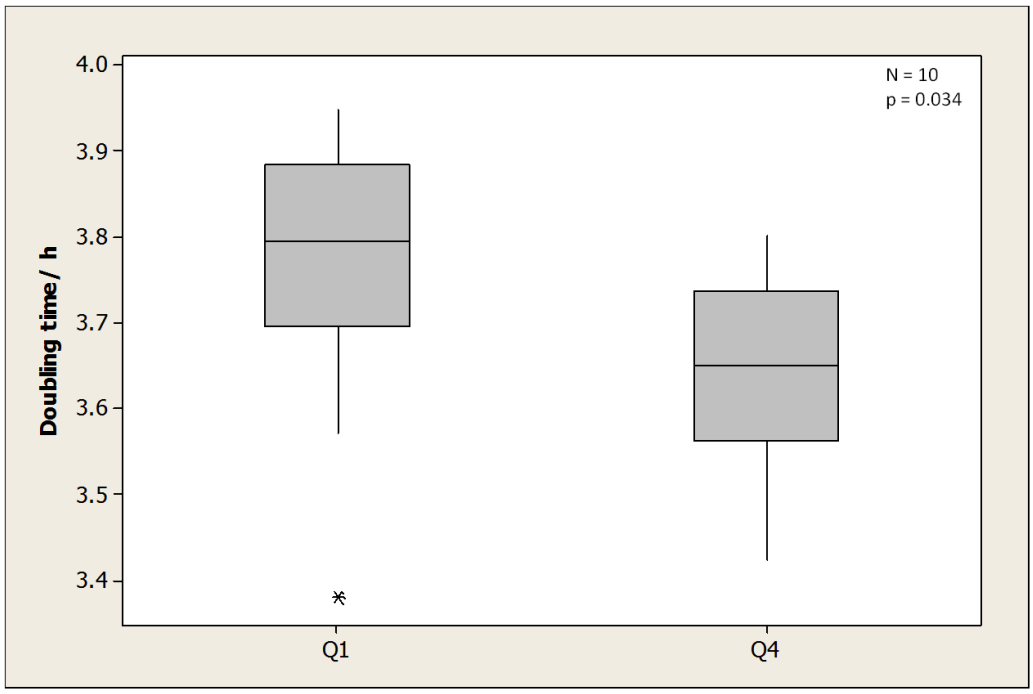
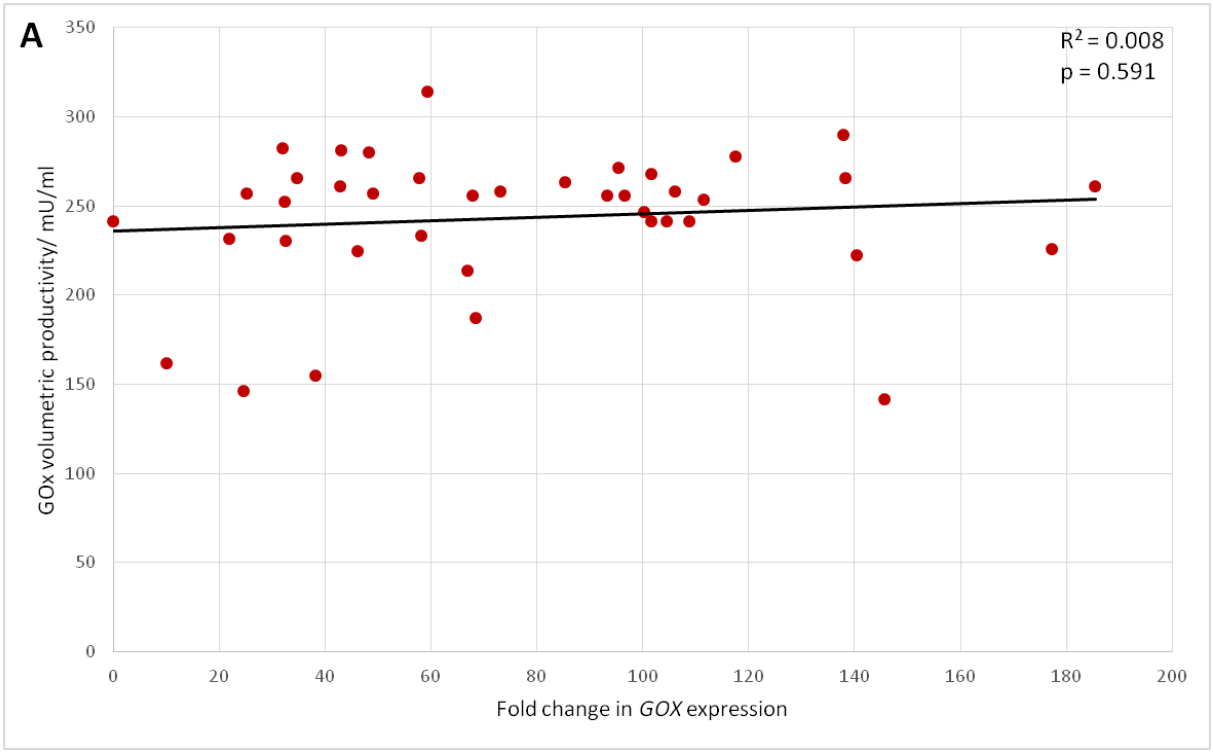


Figure 23: Boxplots comparing the distributions of doubling times for GpaGOxZ clones belonging to Q1 and Q4 for specific productivity from the original test sample. The bottom and top of the boxes represent the interquartile range, and the line within the box represents the median value. The "whiskers" illustrate the 1st and 4th quartiles for the data sets, excluding values outside of 2 standard deviations from the mean (starred). The sample size (N) is displayed and "p" represents the significance value from the Mann-Whitney U test for differences between the median doubling times of each group.

Due to the skewness in the distribution for the doubling times of both Q1 and Q4 the Mann-Whitney U test was deemed more appropriate to test the homogeneity between the 2 groups over the Student's t-test, as it is non-parametric. A significant difference between the median average doubling times within a 95% degree of confidence ($p<0.05$) was observed between clones in Q1 and Q4. Whilst the difference itself appears to be small, with a decrease in 0.143 hours for the median average doubling time in high producing clones, its statistical significance indicates that Q1 and Q4 belong to discrete subpopulations that have diverged in average doubling time from the original cell line. If this is the case, the lower median doubling time suggests that the average growth rate appears to be marginally increased within the subpopulation containing clones with the highest productivities. Whilst the data does not provide an understanding as to whether natural variation in growth provides a certain level of predisposition towards higher protein yields, or if it occurs as a result of clonal variation in productivity, it does suggest that the growth phenotype, comprised of overall growth and growth rate, is a minor predictor for productivity amongst a clonal population.

4.4 Correlating recombinant gene transcription with corresponding protein productivity for GOx

The successful increase in titres of a number of recombinant proteins in response to increasing the number of copies of the corresponding gene within the cell confirms that a link exists between gene dosage and productivity (Cos et al, 2005) (Zhu et al, 2009A) (Norden et al, 2011). However there is little insight as to whether significant variability in expression of recombinant genes at the transcriptional level exists between clones with uniform copy numbers, and if it affects clonal variation. Hypothetically the rate of transcription from the *AOX1* promoter should have a direct relationship with synthesis rates of the encoded, secreted protein if the quantity of transcript available for translation by ER-bound ribosomes is limiting. To analyse the contribution of variation in recombinant gene transcription towards protein productivity within the GpαGOxZ clonal population, RNA was isolated from each clone upon harvest of expression cultures as described in 2.3.17. The expression of *GOX* was quantified by RT-qPCR of the RNA samples using the primer set 40-qgox2-F and 41-qgox2-R and calculated as a fold change in expression relative to the expression of actin (*ACT1*) and plotted against volumetric and specific productivities for the respective clone (Figure 24).



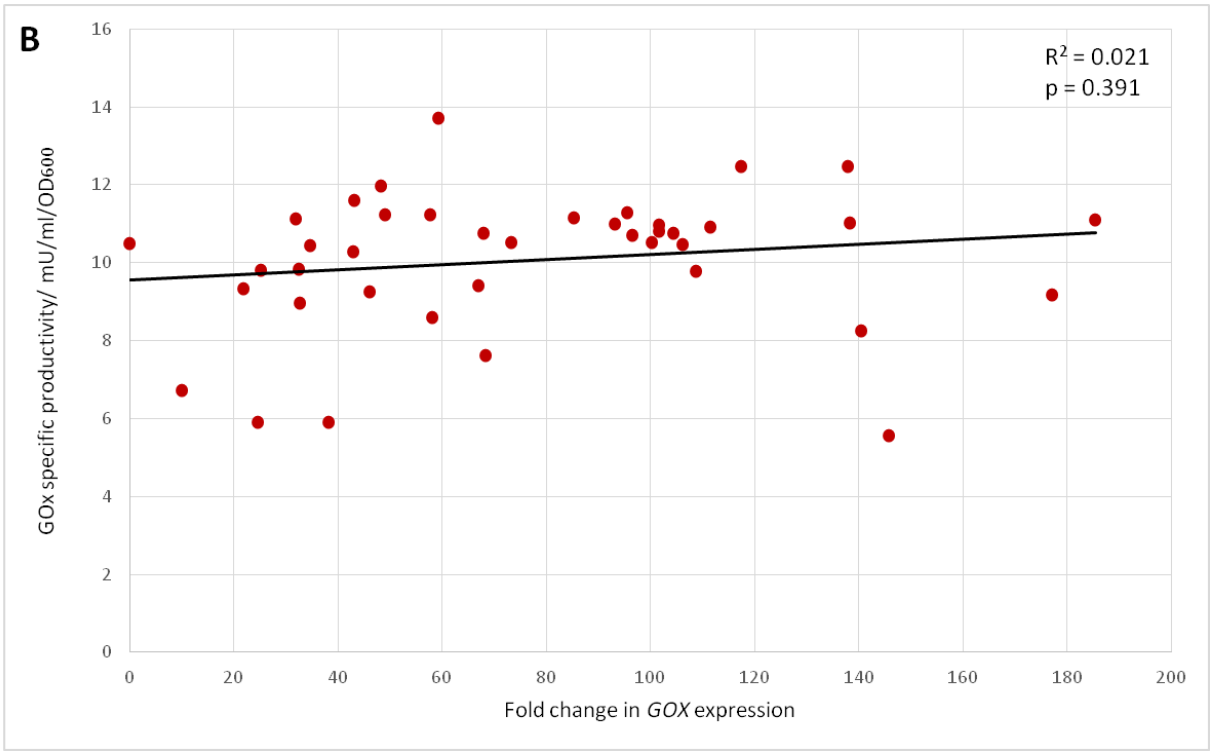


Figure 24: Linear regression between the volumetric productivity (A) or specific productivity (B) of GOx and the expression of recombinant *GOX* relative to the expression of native *ACT1* following a 48 hour expression trial. The R^2 value and significance value (p) from the regression analysis between the variables are displayed.

The range of expression for *GOX* among the sample was surprisingly large and significantly exceeded that of any other genes tested within this project, including those tested using identical cDNA samples. This may however be characteristic of expression of genes from P_{AOX1}, as comparable ranges have previously been derived during methanol-induced expression for alternative recombinant genes (Aw and Leak, 2012). Even with the abnormally high variability in fold expression of *GOX* between clones, no correlation could be found between expression and volumetric/ specific productivity of GOx. This is unexpected because, as surmised previously, a strong increase in transcription would be predicted to increase protein production. A potential explanation could be that expression from P_{AOX1} is prone to naturally high levels of fluctuation between higher and lower states depending on external factors, and that the values calculated for *GOX* expression for each clone may be inappropriate representations of an average rate of transcription. Alternatively, under the set conditions, transcription may not be a limiting factor to productivity, supported by the fact that the minimal copy number drift identified in 3.5 also did not impact GOx production significantly.

4.5 Correlating *HAC1* expression, as a marker for UPR regulation, with GOx productivity

The attribution of bottlenecks within the secretory pathway for *P. pastoris* to recombinant production has led to the examination of ER stress responses and total intracellular protein as indicators of the efficiency of secretion of recombinant proteins (Love et al, 2012). Indeed, the study of the secretion of a library human lysozyme variants ranging in their native state stabilities found that decreasing stability resulted in both lower secreted productivities and an upregulation of UPR and ERAD pathway genes. The implication of the limitations of *P. pastoris*' secretory capacity for productivity has informed the development of high producing strains constitutively expressing certain UPR regulated genes that are beneficial to protein folding, for example the gene encoding protein disulphide isomerise (*PDI*) (Inan et al, 2006). The rationale behind such strains is to artificially increase the folding capacity of the cell's ER, along with the threshold for the activation of the aforementioned stress responses that eventually lead to product loss.

Should natural variation in the ER folding capacity manifest within a clonal population due to unknown environmental or intrinsic factors, it could therefore confer a beneficial adaptation towards

protein secretion that would be partly responsible for the variability observed in the expression system. As previously described (1.6.1.1) *HAC1* encodes the transcriptional regulator for the UPR pathway in eukaryotes and is activated in yeast through the splicing of *HAC1* mRNA to yield a transcript encoding its functional form of the protein. However in *P. pastoris* a second regulatory feature for UPR activation is that the expression of its *HAC1* homologue is also upregulated, indicating its usefulness as an indicator for relative ER stress (Guerfal et al, 2010). Clones found to express lower levels of *HAC1* than others could subsequently be interpreted as experiencing lower levels of ER associated stress in culture and vice versa.

To test for variation in *HAC1* regulation within clonal populations and its relation to productivity, RT-qPCR with *P. pastoris HAC1*-specific primers (32-q*HAC1*-F, 33-q*HAC1*-R) was also carried out on the isolated RNA from the 40 GpαGOxZ clones following expression. Again, the relationship between the 2 variables was examined graphically and through regression analysis (Figure 25).

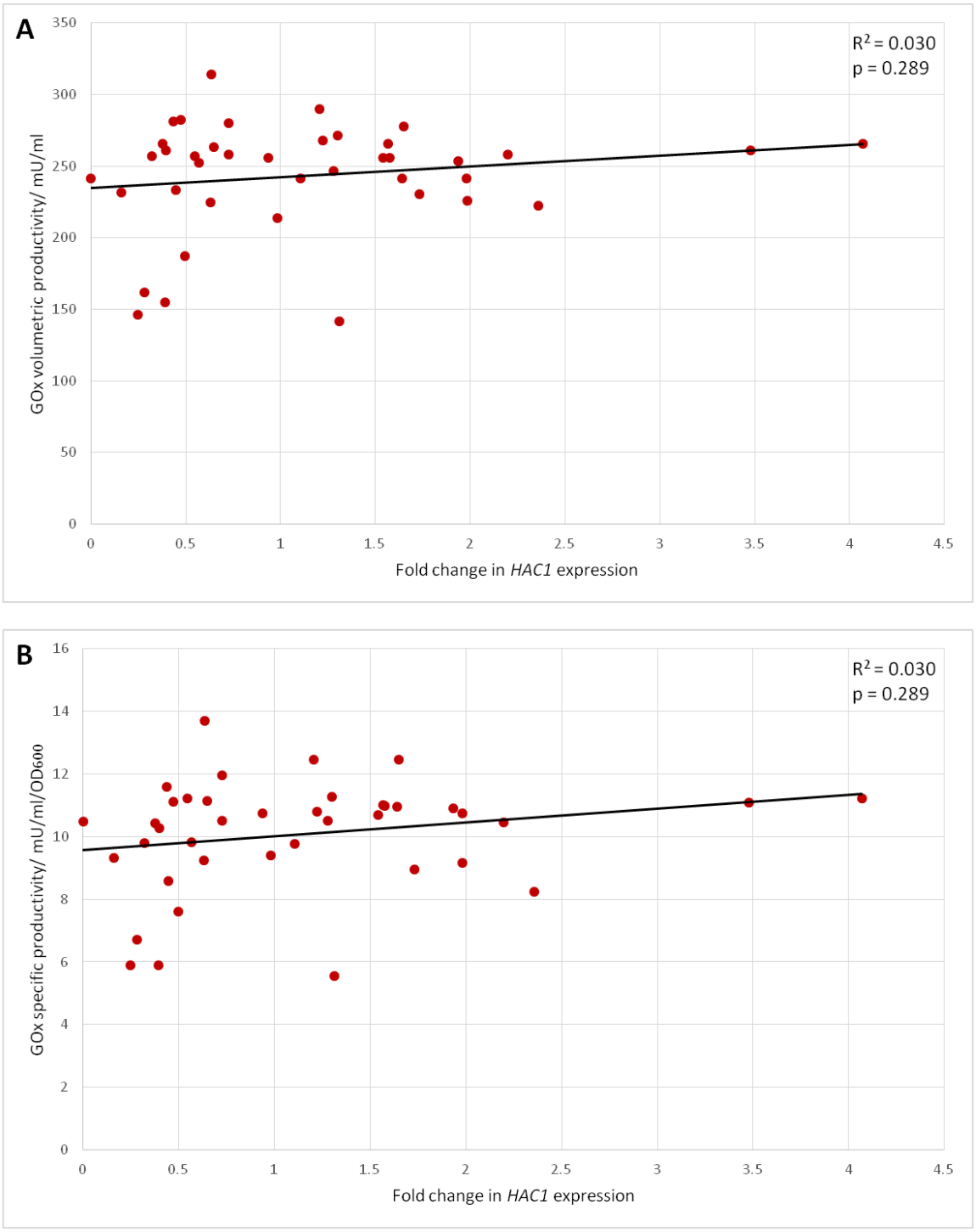


Figure 25: Linear regression between the volumetric productivity (A) or specific productivity (B) of GOx and the expression of native *HAC1* relative to the expression of native ACT1 following a 48 hour expression trial. The R^2 value and significance value (p) from the regression analysis between the variables are displayed.

Compared to *GOX*, the range in expression for *HAC1* within the sample is notably smaller and more typical, with the vast majority of clones clustering between a 0.1 and 1.5 fold change in expression related to *ACT1*. Regression analysis failed to identify any significant correlation between *HAC1* expression and volumetric/ specific productivity of GOx, exemplified by clones expressing relatively high levels of *HAC1* producing roughly equivalent titres to clones with significantly lower expression. The results indicate that impactful variations in ER stress thresholds either do not occur within the clonal population or that they have no effect on productivity for the tested scenario i.e. for the expression of a single copy of *GOX* under the parameters of the expression trial used for the study. It is not clear from this study whether the different levels of *HAC1* expression reflect a natural variation in sensitivity of the ER to ER stress or a natural variability in basal/methanol-induced *HAC1* expression levels.

4.6 Discussion

This study examined the relationship between cell growth, recombinant gene transcription, the transcription of *HAC1* and recombinant protein productivity within a defined sample of clones to evaluate their contribution towards clonal variation. Both the volumetric and specific productivities of GOx were chosen to be investigated as response variables, on the basis that studying both would allow for the differentiation between higher productivities resulting from improved culture growth and higher cell-specific productivity. However, for each of the variables tested, both measures for productivity corroborated each other, providing more valid confirmation for all of the conclusions made from the results.

Of all of the variables tested, only the growth of clones in culture provided a statistically significant correlation with productivity of GOx. However, in the case of the final cell densities recorded for each culture upon harvest, the weakness of the correlation suggests that variation in total cell growth is only a minor predictor of clonal variation amongst other, currently uncharacterised traits. The trade-off between biomass and yield of high value products is a commonly recognised phenomenon across a spectrum of microbial fermentations, in which microorganisms exhibit more inefficient metabolic activity and product formation during rapid growth, instead diverting more energy and resources available in culture towards the formation of cellular mass (Goel et al, 2012). The issue is often

exacerbated for continuous fermentation strategies during which prolonged exposure to the culture environment places selective pressures that favour cells shifting energy towards growth rather than product formation. This could provide an explanation for the negative relationship between final cell density and GOx productivity in the study. Should the antagonistic relationship between total biomass and productivity form the underlying principle for the correlation observed, the strength of the correlation, and therefore the influence of varying total biomass formation on clonal variation, would hypothetically increase as expression is extended for longer periods of time. The divergence towards slightly increased growth rates for the highest producers within the sample could be seen as somewhat contradictory to the inverse relationship between final cell density and productivity, as faster growing cells in culture would also be expected to yield higher cell concentrations in culture. However the difference is small, equating to an approximate difference of 8.4 minutes between the median doubling times of the high and low producers tested. In this case the differentiation in growth rate could indicate minor differences in viability and the prolificacy of individual clones in the culture medium, in which case the findings made by Rebnegger et al (2014) apply. As the study concluded that a higher rate of growth was synonymous with the increased expression of genes involved in protein synthesis and ribosome formation, minor divergences in these expression profiles within clonal populations could potentially be linked to the variable growth and specific productivity observed within this experiment. The higher growth rate of certain clones within a controlled, relatively short expression trial could have benefitted from faster rates of protein synthesis and downregulated stress-related proteolytic responses, resulting in a high productivity phenotype. In this situation protein production and growth rate would increase simultaneously, rather than trading off, utilising substrate from the media at a faster rate than low producers instead of biasing cell growth. Unfortunately cell growth is a complex trait itself, comprised of a number of variables, while also alluding to other relevant phenotypes within cells such as viability and productivity. Therefore it is difficult to ascertain from the data an ideal growth phenotype for recombinant protein production or indeed which aspect of growth is the most beneficial, especially since its variability between clones appears to only have a minor effect on their output. An optimal balance must exist between growth and productivity beyond which one impedes the other and this is often managed in practice through carefully controlling the availability of metabolic substrate during expression as well as screening for suitable clones following the creation of new recombinant strains.

The variation in transcription of *GOX* from P_{AOX1} between clones was unusually high compared to data for other genes, yet did not correlate with the overall productivity of GOx. The AOX1 promoter has

been shown to be tightly controlled and sensitive towards methanol concentrations in culture as well other carbon catabolites (Kupcsulik and Sevella, 2004) (Hartner and Glieder, 2006). For that reason it is likely that the high variability in *GOX* expression could reflect differences between methanol concentrations in each culture due to the variation in growth, and therefore methanol uptake, between clones at the point of harvest. Alternatively, while the on/off states of P_{AOX1} are well regulated, the amplitude of transcription from it could be less so, putting the validity of conclusions made from data collected from P_{AOX1} controlled expression during later stages of methanol induction into question. Nevertheless the results remain unexpected in the context of the strain used within the study, as using a single copy clone as a parent strain should minimise any background variation of recombinant gene copy numbers within the resulting clonal population. It should be surmised that the wide range of GOx expression recorded is not caused by significant copy number variation between clones within the sample but is instead an inherent feature of the AOX1 system. The lack of a positive correlation between *GOX* transcription and GOx titre cannot be fully explained either as a saturation point being reached for *GOX* expression, beyond which increasing transcription no longer has the effect of increasing protein synthesis, as higher producing variants than those observed in this study were found previously (Chapter 3). Another possible justification could be that, considering the high standard rates of transcription achievable from P_{AOX1}, more sizeable increases in transcript quantities are required to bring about notable changes in production of secreted proteins, particularly since previous studies have found its relationship with gene dosage/ copy number is non-linear (Liu et al, 2014) (Aw and Polizzi, 2013). As mentioned previously though, a likely possibility is that the values recorded for GOX expression from P_{AOX1} could have fluctuated over the course of methanol induction and perhaps do not represent an average rate of transcription. In order to test this prediction, and rectify it, a time course experiment for *GOX* transcription throughout various stages of growth, pre and post methanol supplementation would have to be considered in order to ascertain the consistency of P_{AOX1} expression within the parameters of micro expression trials.

No correlation was also found between productivity and the expression of *HAC1*, providing evidence to suggest that variability in UPR induction is not responsible for the clonal variation observed for the expression of a single copy of *GOX*. The average fold change in expression of *HAC1* compared to *ACT1* in the tested clones was 1.17 which is on the lower end of the spectrum for *HAC1* regulation following a 48 hour expression, based on the study of HuL variant secretion by Whyteside et al. (2011), in which all expression values were normalised using the same reference gene as this experiment. In addition to this, the fact that the clonal sample tested was raised from a single copy variant of GpαGOxZ

suggests that the properties and expression rates of GOx in this study do not exert exceptionally high levels of ER stress in cells. Fluctuations in UPR upregulation within cells of GpαGOxZ, possibly in response to shifting rates in secretion throughout expression could not be extensive enough to trigger detrimental effects to productivity through the activation of protein degradation pathways such as ERAD. Based on the links identified between ER stress and secretion rates of recombinant proteins variability in UPR regulation and its threshold for activation might become a contributing factor towards clonal variation for the production of proteins that elicit more critical levels of ER stress. *HAC1*, however, only comprises 1 of many genes that are upregulated during periods of high protein burden on the ER, and so further tests could be conducted to establish the existence of natural variation in a clone's adaptability to secretion bottlenecks and whether it has an effect on its productivity. Factors such as total intracellular protein, expression data for more UPR regulated genes and the regulation of ERAD and other proteolytic pathways could be considered as valid candidates for further investigation as well as the continued search for predictive markers for clonal variation.

5. Deletion of the *OPI1* homologue in *P. pastoris* and its effect on recombinant protein secretion

5.1 Introduction

5.1.1 Aims and objectives

The objective of this experiment was to generate a knockout strains for the *OPI1* homologue in *P. pastoris* and to assess the resulting mutant's capability in secreting a range of recombinant proteins that have been shown to elicit different levels of ER stress. The growth and UPR profiles of strains undergoing methanol-induced expression were also compared to the native *P. pastoris* expression system, and the study aimed to quantify changes in ER membrane size in *Δopi1* strains to determine whether the function of Opi1p in *P. pastoris* is conserved from *S. cerevisae*. The results of the study could be used to establish whether the loss of Opi1p confers an adaptation to the higher protein burden artificially introduced by recombinant protein overproduction and improves cell-specific productivity.

5.1.2 Constitutive expression of UPR elements as a strategy to alleviate secretion bottlenecks

The protein processing capacity of the ER has been identified as a limiting step impeding the secretion of recombinant proteins in *P. pastoris*, as large fractions of total protein targeted to the secretory pathway remain located intracellularly during *AOX1* – based expression (Love et al, 2012). The increase in unfolded protein concentration within the ER lumen induces upregulation of expression from promoters containing unfolded protein response elements (UPREs), signifying the activation of UPR to address the saturation of protein processing and folding machinery in the ER (Gasser et al, 2007A). Since continual upregulation of UPR is intrinsically linked to the triggering of ERAD (Friedlander et al, 2000; Travers et al, 2000) it is unsurprising that its excessive activation is correlated with declines in overall productivity; the phenomenon is exacerbated in strains secreting proteins

with lower native state stabilities (Kumita et al, 2006) or in certain multi-copy strains (Zhu et al, 2009B).The processing limitations of the ER and the subsequent activation of stress responses therefore place constraints not only on the final protein titres achievable using the *P. pastoris* system, but also the organism's amenability to producing adequate yields of complex proteins that require more post-translational modifications or are more prone to misfolding (Ahmad et al, 2014). However the molecular processes within the ER governed by the UPR are in fact beneficial to protein processing and trafficking. The transcription of genes encoding protein disulphide isomerase (Pdi), the molecular chaperone Kar2p and a facilitator of oxidative folding (Ero1) was increased significantly in *P. pastoris* mutants overexpressing the UPR regulator *HAC1* (Gasser, 2007A). Each of these proteins are classed as UPR-regulated factors in *S. cerevisae* and are known to facilitate protein folding and protein disulphide formation in the ER (Normington et al, 1989; Frand and Kaiser 1998; Xiao et al, 2004; Zito, 2015). UPR factors that perform post-translational modifications and assist in protein folding have previously been identified as potential candidates for reducing secretion bottlenecks in *P. pastoris* with the hypothesis that uncoupling their expression from UPR regulation will have the effect of permanently increasing the protein folding capacity of the ER without activating ER stress responses. This strategy has been successful in increasing productivity of certain proteins such as *Necator americanus* secretory protein (Na-ASP1) in a recombinant strain expressing *PDI* under the control of the *AOX1* promoter (Inan et al, 2006). It is possible that overexpression of *PDI* was especially advantageous to the production of Na-ASP1 as preliminary structural studies of the protein predicted it to contain 10 disulphide bonds, and could explain how strains containing 8 copies of the *PDI* overexpression cassette were able to increase final titres up to ~4.5 - fold in strains containing 2 copies of the Na-ASP1 gene (Inan et al, 2006). Expression of the *S. cerevisae PDI* under the constitutive *GAP* promoter was also able to increase the production of the 2F5 anti-HIV Fab fragment by an average of 1.9 fold, which was otherwise accumulating at high levels intracellularly in the native host (Gasser et al, 2005). The same study was even able to increase productivity through the *GAP*-regulated overexpression of *HAC1*, although not as well as for *PDI* overexpression. Again examining the secretion of 2F5 Fab, strains overexpressing 15 different genes involved in ER stress or the secretory pathway, and found to be transcriptionally upregulated during recombinant protein production, were screened for their effect on volumetric productivity (Gasser et al, 2007B). The screen was able to identify previously untested UPR factors and chaperones that were able to increase 2F5 Fab productivity when expressed under the *GAP* promoter, including Kar2p, Ero1, and 2 other chaperones – Ssa4 and Sse1. Overexpression of *KAR2* under P_{AOX1} was also able to increase the

secreted productivity of a recombinant A33 single- chain antibody fragment by up to 3-fold of the control (Damasceno et al, 2007). However no change in product yield was observed when *PDI* was overexpressed, suggesting that the efficacy of overexpression of different UPREs is dependent on the properties of the secreted product (Damasceno et al, 2007). In spite of this, the uncoupling of UPR factors from their native regulation, through expression under high strength promoters, has emerged as a valid approach for increasing recombinant protein titres and warrants the consideration of other UPR-specific pathways that act to increase the processing capacity of the ER as potential candidates for strain improvement.

5.1.3 The role of ER membrane expansion in UPR and protein trafficking

While the precise mechanism of its action is yet to be completely understood, the activation of UPR alters the homeostasis of the ER in yeast and mammalian cells (Shaffer et al, 2004; Schuck et al, 2009). The result is an increase in ER – resident phospholipid biogenesis and an overall expansion in ER volume. In *S. cerevisae* the ER membrane contains a high phosphatidylinositol content, the biosynthesis of which requires inositol as a precursor and is regulated by a heterodimeric transcription factor complex comprising the proteins Ino2p and Ino4p (Ambroziak and Henry, 1994) (Block – Alper et al, 2002). The native function of these 2 proteins, as well as UPR regulation is essential for correct ER membrane expansion, as the deletion of *INO2* or *HAC1* results in cells that either form aberrant, considerably smaller ER structures (Block-Alper et al, 2002) or fail to increase ER size in response to exposure to ER stressors (Schuck et al, 2009) respectively. Membrane expansion during UPR also appears to be conserved in *P. pastoris*, where cells constitutively expressing *HAC1* were found to form highly ordered, stacked intracellular membrane structures that are absent in un – stressed cells (Guerfal et al, 2010). The actual physiological role of ER membrane expansion in response to excessive ER stress remains to be completely understood, although it is suggested that it acts to increase the ER's processing capacity by providing larger lumenal spaces to house a higher secretory protein content (Cox, Chapman and Walter, 1997; Schuck et al, 2009). This would have the benefit of reducing molecular crowding within the ER, theoretically decreasing the probability of misfolding proteins coming into contact to form insoluble aggregates, as well as providing larger areas of ER membrane to accommodate more ER membrane resident proteins that assist in folding. Also worth noting is that ER expansion is often observed as an adaptation within specialised cells that

secrete large quantities of proteins in nature. Examples include human B lymphocytes, which exhibit an increase in ER volume by approximately 3.7 fold during their differentiation into plasma cells secreting antibodies (Wiest et al, 1990), and pancreatic exocrine cells, known to secrete large quantities of enzymatic proteins and also contain large, stacked rough ERs (Lodish et al, 2000).

5.1.4 Opi1p is a negative regulator of lipid biogenesis and ER membrane size in *S. cerevisae*

Opi1p is classed as a leucine zipper transcriptional repressor of lipid biogenesis genes in yeast (White and Henry 1991), although the majority of characterisation studies have only been conducted on the *S. cerevisae* homologue. In *S. cerevisae* Opi1p was found to repress inositol and phopholipid biogenesis pathways by sequestering Ino2p/Ino4p from the upstream activating sequences on promoters of Ino2/Ino4 – regulated genes (UAS_{INO}) when free inositol levels are abundant in the cell. Ino2p/Ino4p complexes bound with Opi1p fail to activate the transcription of inositol-responsive phospholipid biogenesis genes such as *INO1* (Wagner and Dietz, 2001). The repressive effects of Opi1p are reversed under conditions where exogenous levels of inositol are low, and Opi1p is sequestered to the ER through associating with the ER bound protein Scs2p, where it can no longer inhibit transcription (loewen, 2004). The recruitment of Opi1p by Scs2p also requires functional Hac1p, suggesting that the regulation of Opi1p forms part of the molecular switch for activating phospholipid biogenesis and ER expansion during UPR (Brickner and Walter, 2004). This was supported further through deletion studies for *OPI*, which found that *Δopi1* mutants contain a constitutively expanded ER during growth on standard YPD medium (Schuck et al, 2009). Furthermore the role of ER expansion in alleviating ER associated stress was confirmed through the complementation of UPR – deficient mutants with the deletion of *OPI1* (Schuck et al, 2009). Mutants lacking the *HAC1* gene, as used in the study, are unable to initiate UPR and do not grow in the presence of the artificial ER chemical stressor tunicamycin. The deletion of *OPI* partially restored cell viability of *Δhac1* strains exposed to 0.05µg/ml tunicamycin, suggesting that *Δopi1* mutants exhibit an independent form of ER stress alleviation to the UPR through separating ER expansion from UPR regulation (Schuck et al, 2009). The disruption of *OPI1* could therefore have positive implications for the improvement of expression strains as it uncouples a process that increases the processing capabilities of the ER, in this case ER expansion, from UPR regulation. This has been recently verified

in expression trials of *S. cerevisae* secreting recombinant Immunoglobulin-G (IgG), the results of which demonstrated that expression with *Δopi1* mutants increased productivity up to 4 fold (de Ruijter, Koskela and Frey, 2016).

5.2 Identification and deletion of the *P. pastoris OPI1* homologue

A local protein sequence alignment with BLASTp (http://blast.ncbi.nlm.nih.gov) revealed an uncharacterised 388aa protein in *P. pastoris* with sequence similarity to Opi1p in *S. cerevisae* S288C. Although the putative protein only shares 27% identity, spanning 87% of *S. cerevisae* S288C Opi1p, further alignments against the sequence returned a 25-45% identity to a number of proteins classed as Opi1p homologues in other yeast species including *Kluyveromyces marxianus*, *Candida utilis* and *Candida glabrata*. The conserved functional domain containing a leucine zipper motif and belonging to the Opi1p superfamily was also identified between amino acid residues 249 and 359, supporting the likelihood of the protein being a putative Opi1p homologue in *P. pastoris* (and is henceforth referred to as Opi1p).

The adjacent sequence located immediately 5' and 3' to the corresponding gene within the NRRL 11430 genome was obtained from the NCBI database (http://www.ncbi.nlm.nih.gov/) and used to design a knockout construct for *P. pastoris OPI1*. Since the removal of only a single gene was required, the targeted gene replacement of *OPI1* with an antibiotic resistance cassette was selected as the method for generating *Δopi1* mutants (Figure 26).

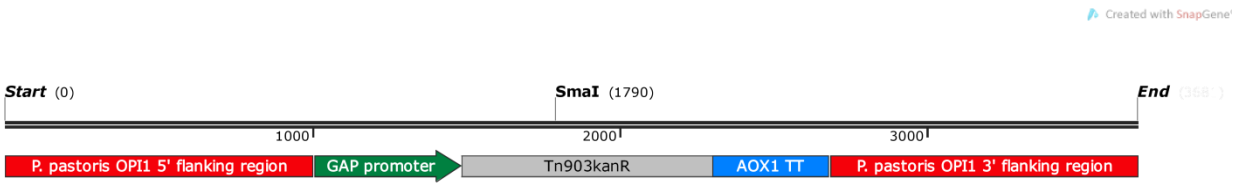


Figure 26: Map of the 3681bp linear knockout construct for the targeted replacement of *OPI1* in *P. pastoris* NRRL-11430 with a G418 resistance cassette. The SmaI restriction site is present in *tn903kanr* but absent in *OPI1*, enabling initial confirmation of *OPI1* removal by restriction digest.

The knockout construct contains the *tn903kanr* gene expressed under the constitutive *GAP* promoter, conferring resistance to G418 in yeast. The inclusion of a $G418^R$ cassette as a selectable marker allows the resulting *Δopi1* mutants to remain compatible with the integration of pPICZα and pAVE522 – based expression vectors utilising the zeocin resistance marker - *sh ble* for selection. The cassette is

flanked by sequences identical to the 1kb regions located immediately 5' and 3' to *OPI1* within the NRRL 11430 genome to stimulate the replacement of *OPI1* through a double crossover event, mediated by homologous recombination.

The 5' flanking region of *OPI1* was PCR amplified from NRRL 11430 genomic DNA with the primers 18-Opi1kan5'-F and 19-Opi1kan5'-R, whereas the 3' region was amplified with 22-Opi1kan 3'-F and 23-Opi1kan 3'-R (9.1). The primers 19-Opi1kan5'-R and 22-Opi1kan 3'-F contained complementary ends to the 5' and 3' ends of the $G418^R$ cassette respectively. The $G418^R$ cassette, supplied as a custom DNA fragment by Fujifilm Diosynth Biotechnologies (Billingham, UK), was PCR amplified to contain complementary ends to the 5' and 3' *OPI1* flanking regions with the primers 20-Opi1kan-F and 21-Opi1kan-R (9.1). The 3 PCR amplicons were gel purified and joined by Gibson assembly. To circumvent the requirement for cloning into *E. coli*, the reaction mix containing the fully assembled knockout construct was used as a template for PCR with the primers 18-Opi1kan5'-F and 23-Opi1kan 3'-R to generate sufficient quantities of the construct for direct transformation into *P. pastoris*. DNA fragments of the size corresponding to the full knockout construct were gel purified and cloned into *P. pastoris* NRRL 11430 and the recombinant strain - CLD804 by electroporation, with selection on 0.5mg/ml G418 sulphate.

A subsequent PCR analysis of genomic DNA isolated from NRRL 11430 and CLD804 colonies transformed with the *OPI1* knockout construct was conducted with the primers 26-Opi1-F and 27-Opi1-R, which specifically anneal to *OPI*. The PCR revealed a high number of false positives that were resistant to 0.5mg/ml G418 sulphate but retained the *OPI1* gene (Figure 27).

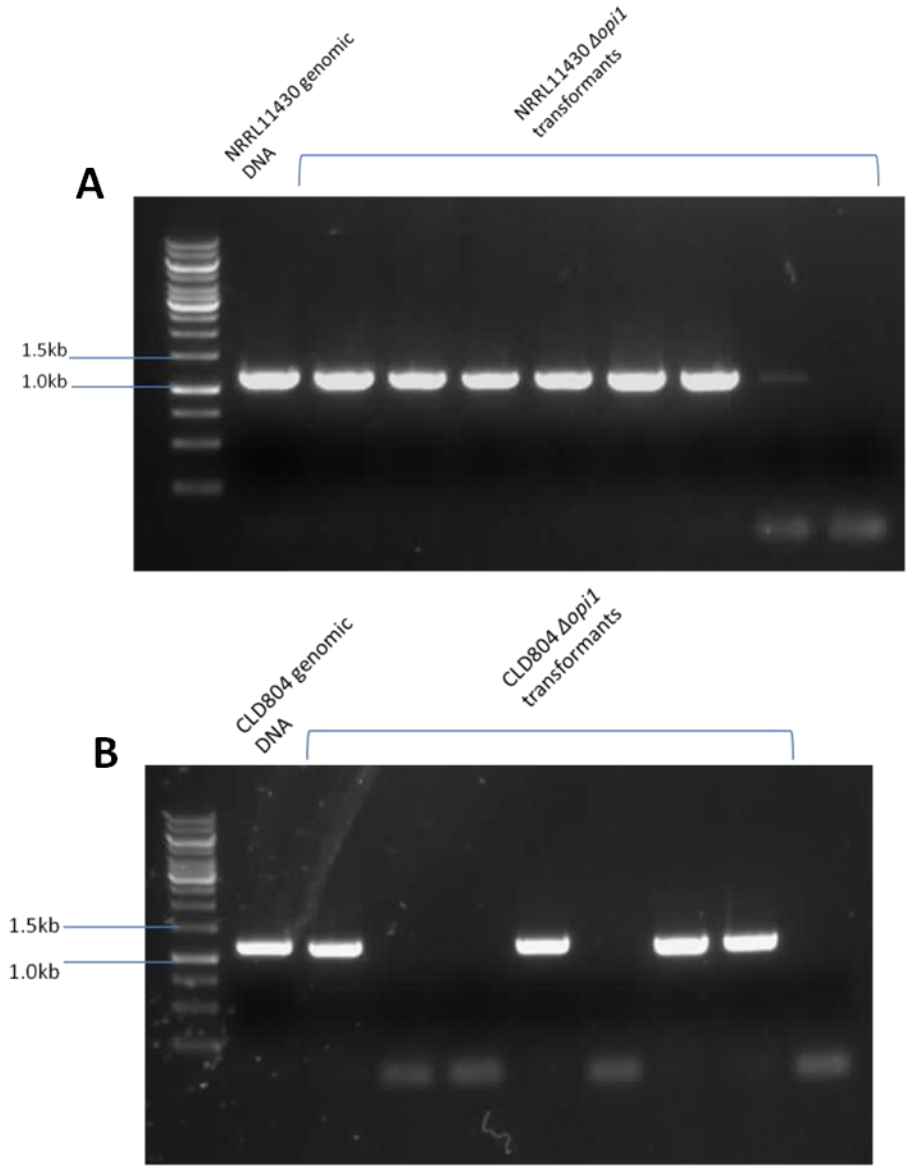


Figure 27: PCR screen of genomic DNA from clones following transformation with the *OPI1* knockout construct, using primers specific to *OPI1*. A – clones from *P. pastoris* NRRL-11430 and B – clones from *P. pastoris* CLD804. The PCR of genomic DNA from the native strain is included in each case as a negative control.

This could have been caused by transient expression of *tn903kanr* from unintegrated expression cassettes within cells, or the insertion of the knockout construct into the *OPI1* locus as a single crossover event without removing *OPI1*. Even so the PCR screen identified a number of clones that had potentially lost *OPI1*. Further testing to confirm the removal of *OPI1* was conducted by PCR amplifying a section of the *OPI1* locus, spanning the location of *OPI1* and flanking sequence, from the genomic DNA of potential positive clones using the primers 28-Opi15'flankseq and 29-Opi13'flankseq (9.2). A single SmaI restriction site is located centrally within the *OPI1* knockout construct but is not found in the *OPI1* gene. Sequencing of the PCR amplified *OPI1* loci with Opi15'flankseq and 29-Opi13'flankseq, complemented by a restriction digest of the loci with SmaI confirmed the replacement of *OPI1* with the G418^R cassette within the tested NRRL 11430 and CLD804 clones (Figure 28).

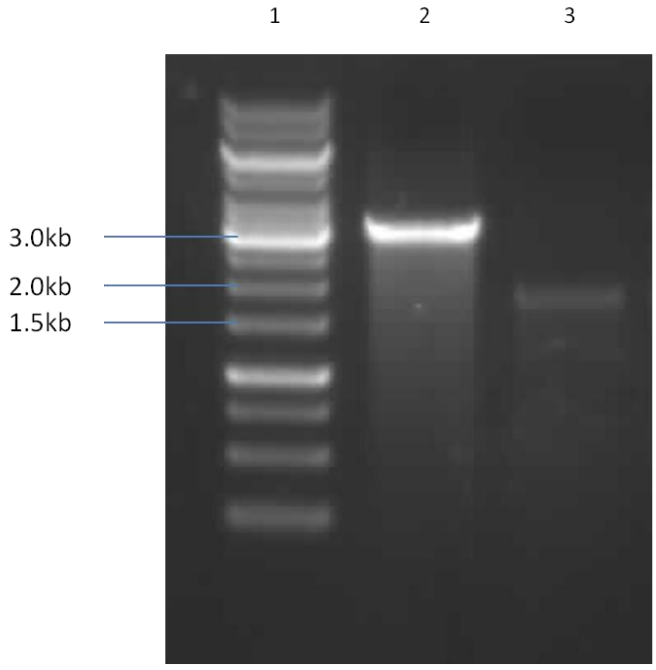


Figure 28: Digest of PCR amplified OPI1 locus DNA from NRRL 11430 Δopi1 with SmaI. The presence of the smaller ~1.8kb bands confirms the replacement of OPI1 with the knockout cassette. Lane 1 - 1kb DNA size marker, 2 - OPI1 locus from NRRL 11430 cut with Sma1, 3 - OPI1 locus from NRRL 11430 Δopi1 cut with SmaI.

5.3 Growth curve analysis and UPR regulation of *Δopi1* mutants

Since the deletion of *OPI1* would potentially derepress a number of biosynthetic pathways, it is possible that the growth of *Δopi1* mutants would be affected under certain conditions as the carbon flux into cells is unevenly distributed away from essential metabolic pathways and towards formerly Opi1p regulated processes. To assess whether the loss of *OPI1* is detrimental to cell viability when cells are cultured in standard growth and expression media, growth curves were compared between NRRL 11430 and NRRL 11430 *Δopi1* in BMG and BMM media, containing 1% (v/v) glycerol and 0.5% (v/v) methanol respectively. 3 biological replicates for each strain were initially grown for ~16 hours in 5ml BMG at 30°C, 250rpm before subculturing into 25ml of the appropriate medium in 250ml baffled flasks at a normalised, starting OD_{600} of 0.1. During exponential phase the cell densities were recorded at 1 hour intervals until the stationary phase of growth had been reached in each culture and plotted (Figure 29, 30).

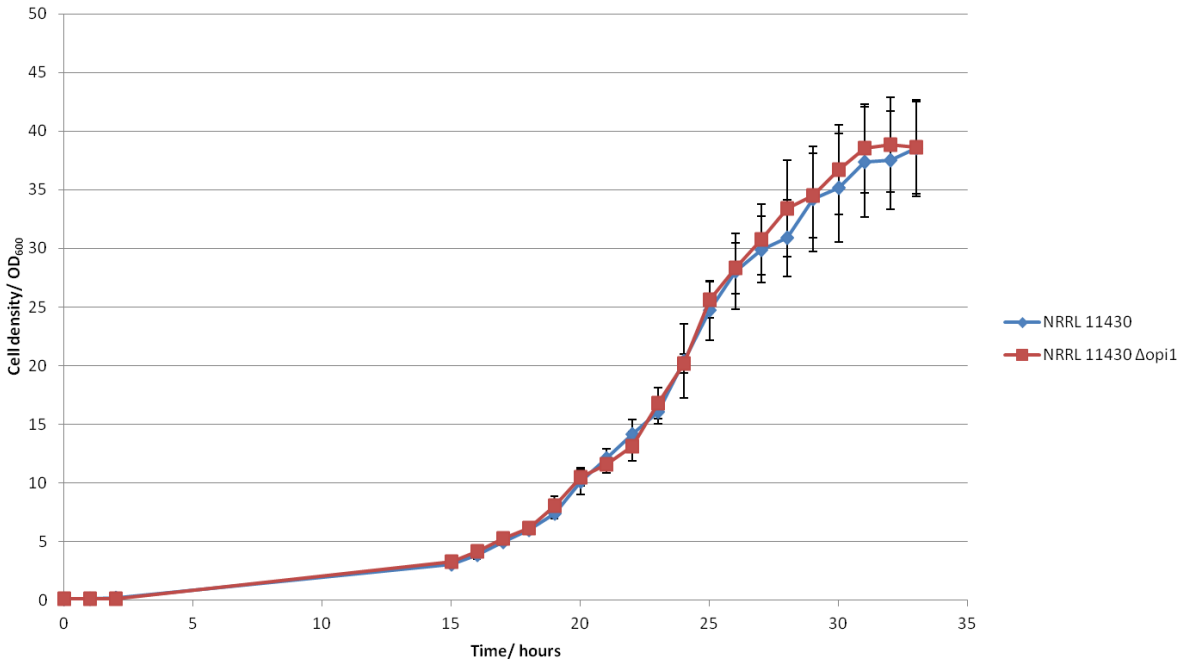


Figure 29: Growth curve of NRRL-11430 and NRRL-11430 *Δopi1* on BMG media containing 1% (v/v) glycerol. Error bars represent the standard deviation of 3 biological replicates.

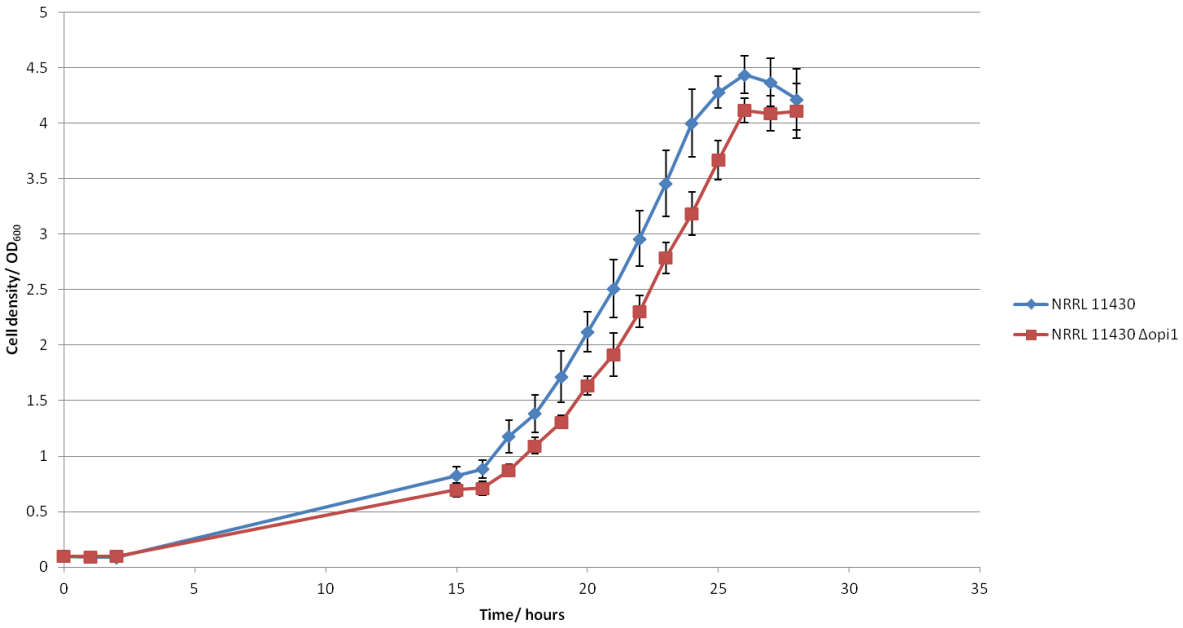


Figure 30: Growth curve of NRRL-11430 and NRRL-11430 Δopi1 on BMM media containing 0.5% (v/v) methanol. Error bars represent the standard deviation of 3 biological replicates.

The growth characteristics of NRRL 11430 *Δopi* were observed to be similar to native NRRL 11430 when cultured in 1% (v/v) glycerol. No notable changes in the growth rate during exponential phase or total growth occurred, suggesting that *Δopi* mutants do not experience reduced growth on typical concentrations of glycerol used in yeast media. A slight shift in growth was observed for *Δopi* cells cultured in methanol, although no differences in final cell density were found. The results from the growth curve may indicate that the duration of lag phase growth was increased in NRRL 11430 *Δopi* compared to NRRL 11430 but, once cells entered the exponential phase, that growth rate and total growth were indistinguishable from the wild type. Most of the growth plotted for methanol appears to be linear rather than logarithmic, suggesting the incidence of oxygen limitation in the flask cultures (Elsworth, Williams and Harris-Smith, 1957) (Fredlund et al, 2004). This could then potentially mask other differences in the growth phenotype of the two strains.

To monitor the basal levels of UPR activation in NRRL 11430 *Δopi* under conditions that do not induce ER stress and under growth conditions experienced during methanol induction, the expression of *HAC1* and the UPR factors *PDI* and *KAR2* were measured by RT-qPCR for cells cultured in YPD and BMMY media. Single colonies of NRRL 11430 and NRRL 11430 *Δopi* were grown for ~16 hours in 5ml of either YPD or BMMY media until exponential growth was reached, after which RNA was isolated from 1ml of culture (Figure 31, 32). The relative expression levels of *HAC1*, *PDI* and *KAR2* were then quantified for each RNA sample by RT-qPCR with the primer sets specified for each gene in section 2.3.4.1.

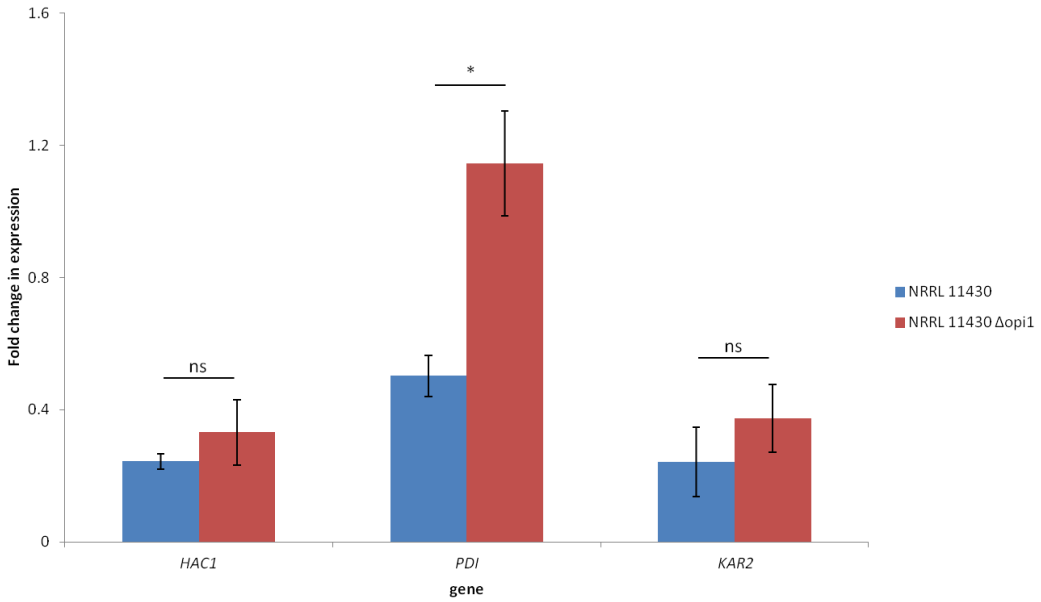


Figure 31: Expression of HAC1, PDI and KAR2 by NRRL 11430 and NRRL 11430 Δopi1 during exponential growth in YPD media. Values are given as a fold change in expression relative to ACT1. Error bars represent the standard error of 3 biological replicates. The significance levels of unpaired t - tests between strains for the expression of each gene are displayed – ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001

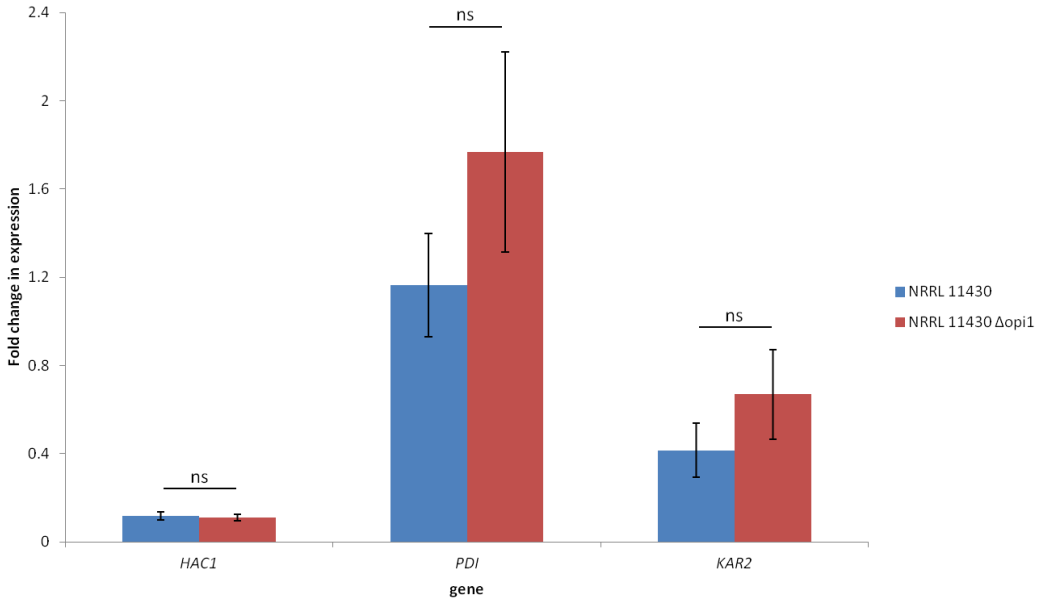


Figure 32: Expression of HAC1, PDI and KAR2 by NRRL 11430 and NRRL 11430 Δopi1 during exponential growth in BMMY media (0.5% (v/v) methanol). Values are given as a fold change in expression relative to ACT1. Error bars represent the standard error of 3 biological replicates. The significance levels of unpaired t - tests between strains for the expression of each gene are displayed – ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001

As expected during growth on a complete medium such as YPD, with no exposure to artificial ER stressors, the expression levels of the UPR regulated genes *HAC1*, *PDI* and *KAR2* remained low in NRRL 11430 in comparison to values observed within this project and previous studies. However *PDI* was comparatively upregulated in NRRL 11430 *Δopi* to ~2.1 fold of expression in NRRL 11430. A similar pattern was observed when cells were grown in BMMY media, although the mean increase in *PDI* expression in NRRL 11430 *Δopi* was found to be statistically insignificant for the data set provided by the experiment. The increased expression of *PDI* would normally signify a cellular response to increased levels of proteins requiring disulphide bond formation in the ER. The results suggest either that protein traffic through the ER could be increased in *Δopi* mutants, or that the transcriptional regulation of *PDI* specifically could have been altered by the loss of *OPI1*, since the lack of a significant increase in *HAC1* or *KAR2* expression implies that conventional UPR has not been activated in NRRL 11430 *Δopi.*

5.4 Specific productivity and induction of the UPR during recombinant protein secretion in *Δopi1* mutants

5.4.1 Aprotinin production in CLD804 *Δopi1*

As a preliminary analysis of the effect of the loss of OPI1 on the production of secreted, recombinant proteins using the AOX1 system, the specific productivity of CLD804 *Δopi1* expressing aprotinin was compared to that of native CLD804. Flask expressions for each strain were conducted by normalising the OD600 of cultures to 1.0 in 25ml BMMY, and expressing over 5 days with the continued addition of 0.5% (v/v) methanol. The quantity of secreted aprotinin was assayed by the densitometric analysis of SDS – PAGE for each culture supernatant and used to calculate the mean specific productivity of CLD804 and CLD804 *Δopi1* (Figure 33).

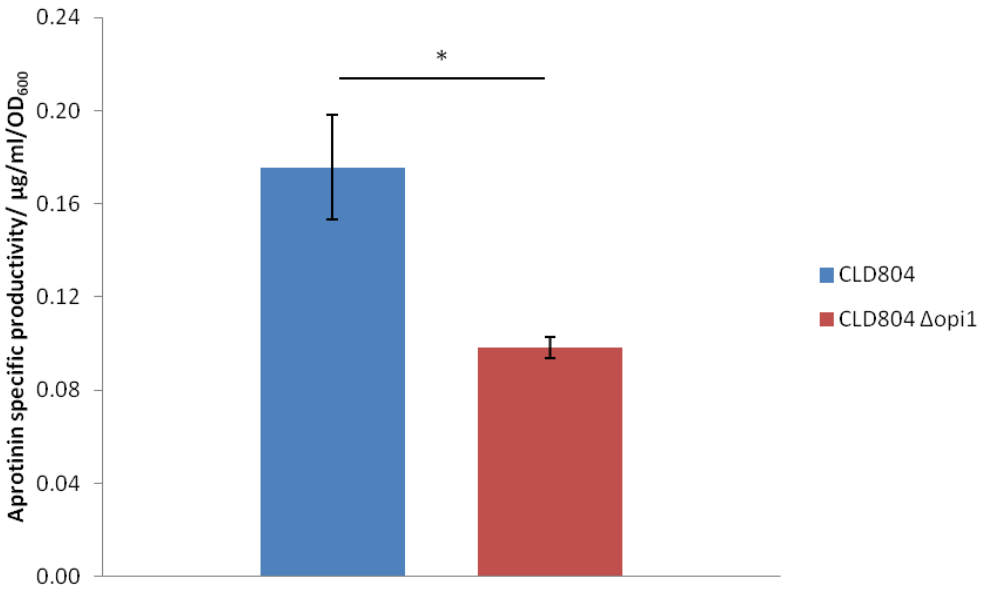


Figure 33: Mean specific productivity of aprotinin for *P. pastoris* CLD804 and *P. pastoris* CLD804 *Δopi1* following expression in BMMY over 5 days at 28°C. Error bars represent the standard error of 5 biological replicates.

Surprisingly the specific productivity of secreted aprotinin was significantly reduced in CLD804 *Δopi1* to approximately 55% of the aprotinin titre produced by CLD804. The CLD804 strain was originally generated by screening large samples of transformants for high productivity following multiple rounds of selection with PTVA, and therefore represents a near – optimal productivity phenotype for aprotinin. It could therefore be possible that, during the growth and transformation of CLD804 with the OPI1 knockout cassette that a reduction in optimal productivity unrelated to the deletion of OPI1 could have occurred through mechanisms such as copy number loss by loop out recombination (Aw and Polizzi, 2013) or a drift in the production phenotype through sequential generations of daughter cells (Love et al, 2010). Nevertheless the initial results raise concerns over the productivity of *Δopi1* mutants and will require further, detailed comparisons between more equivalent recombinant strains.

5.4.2 Specific productivity and UPR induction of *Δopi1* mutants secreting glucose oxidase

To account for secondary factors that could affect productivity such as copy number variation of the target gene and the background of clonal variation between initial transformants, both native NRRL 11430 and NRRL 11430 *Δopi1* were transformed with pGOxZα, assembled previously and containing the *P. funiculosum* glucose oxidase gene under control of P_{AOX1}. The resulting transformants were immediately screened for the copy number of *GOX* by qPCR of genomic DNA samples with the primers 40-qGOX2-F and 41-qGOX2-R in order to isolate 5 single copy integrants of pGOxZα for both NRRL 11430 and NRRL 11430 *Δopi1*. The strains, renamed NpαGOxZ and NpαGOxZ *Δopi1* were induced to express GOX in shake flask cultures of 25ml BMMY from a starting OD_{600} of 1.0 over a 48 hour period. The final GOx titre in the supernatant of each culture was then assayed to calculate the mean specific productivity for each strain. Finally, the expression of *HAC1*, *PDI* and *KAR2* relative to their expression in NRRL 11430 under identical conditions was measured by RT-qPCR of RNA samples isolated post –harvest to quantify the state of UPR activation in NpαGOxZ and NpαGOxZ *Δopi1* expressing single copies of *GOX* (Figure 34).

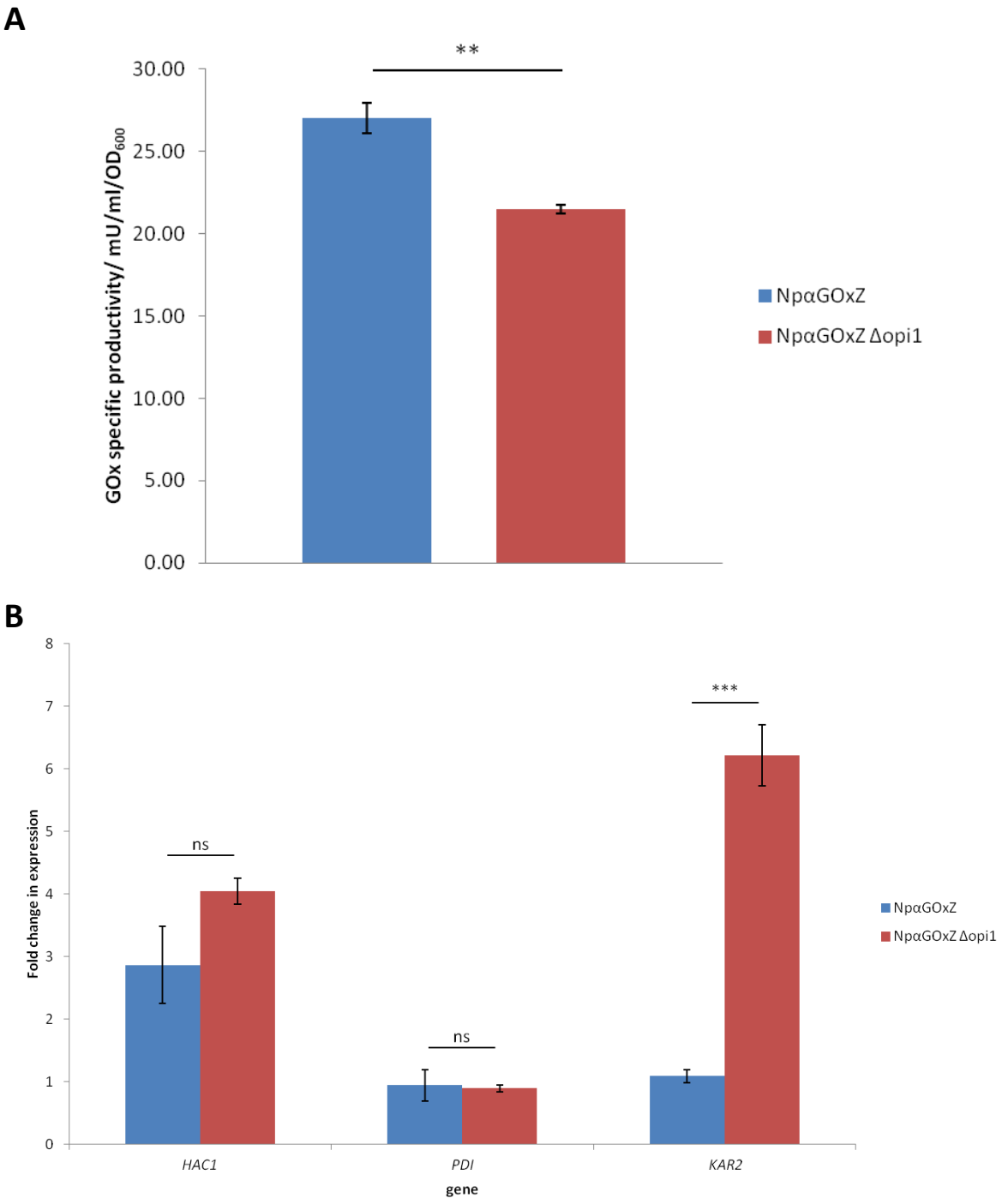


Figure 34: Mean GOx titres (A) and fold change in expression of the UPR genes *HAC1*, *PDI* and *KAR2* relative to wild type NRRL 11430 levels (B) of NpαGOxZ and NpαGOxZ *Δopi1* following a 48 hour expression trial. Error bars represent the standard error of 5 biological replicates and the significance levels of unpaired t-tests between strains for the expression of each gene are displayed – ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001

A decrease in the mean specific productivity was again observed in the *Δopi1* mutant, although the difference was less dramatic than for the production of aprotinin, as NpαGOxZ *Δopi1* produced approximately 20% less GOx than NpαGOxZ. The most notable difference was observed in the regulation of UPR genes in NpαGOxZ *Δopi1* following methanol induction over 48 hours, specifically for *KAR2* – which was upregulated to ~6 – fold of its expression in NpαGOxZ under identical conditions. The mean expression of *HAC1* was also increased in NpαGOxZ *Δopi1* over NpαGOxZ although the difference was not found to be significant; most likely due to the higher degree of variation in *HAC1* expression between the biological replicates of NpαGOxZ. Although *HAC1* expression was upregulated by nearly 3 fold, native NpαGOxZ exhibited a negligible difference in the expression of *KAR2* and *PDI* from wild type *P. pastoris*, suggesting that the expression of a single copy of *GOX* over 48 hours should not elicit excessive levels of ER stress. Therefore considering the large increase in *KAR2* expression together with the overall decrease in secreted productivity per cell; the results indicate that *Δopi1* cells expressing *GOX* at a similar transcriptional rate to the native strain experience greater ER stress and a higher UPR activation state for a theoretically identical influx of unfolded proteins into the secretory system.

5.4.3 Specific productivity and UPR induction of *Δopi1* mutants secreting synthetic human lysozyme T70A

Since the deletion of *OPI1* in *S. cerevisae* uncoupled the expansion of the ER from the UPR, it was hypothesised that *Δopi1* mutants would produce ERs with higher protein folding and processing capacities during unstressed conditions. Therefore the *Δopi1* phenotype would provide an advantage specifically in the secretion of recombinant proteins that would normally elicit high levels of UPR in cells. To evaluate the performance of *P. pastoris Δopi1* under conditions inducing a strong activation of UPR during recombinant protein production, control and *Δopi1* – based strains expressing synthetic human lysozyme T70A (HuL T70A) were generated. The T70A mutation of HuL reduces the native state stability of the protein's tertiary structure, increasing its propensity to misfold and accumulate within the ER in an aggregated state (Kumita et al, 2006). Consequently the secretion of over expressed *HuL T70A* by *P. pastoris* has been demonstrated to stimulate high levels of UPR, resulting in increases for the transcription of all 3 of the UPR genes tested in this study - *HAC1*, *PDI* and *KAR2*

(Whyteside et al, 2011). The expression of *HUL T70A* under P_{AOX1} has hence been established as an effective molecular tool for inducing UPR through recombinant protein secretion.

To generate strains expressing *HUL T70A*, NRRL 11430 and NRRL 11430 *Δopi1* were transformed with the vector pPICzm-T70A, containing *HUL T70A* within an expression cassette comprising P_{AOX1} and the *S. cerevisae* α-mating factor secretion signal. For ease of identification the new recombinant strains were designated NAαT70A and NAαT70A *Δopi1*. Again, to account for differences in productivity occurring due to copy number variation, transformants were screened for the number of integrated copies of *HUL T70A* by qPCR of genomic DNA with the primers 83-qsynHuL3-F and 84-qsynHuL3-R. Once 5 single copy integrants had been identified, the strains were induced to express *HUL T70A* in shake flask cultures of 25ml BMMY from a starting OD_{600} of 1.0 over a 48 hour period. The final HuL T70A titre in the supernatant of each culture was then quantified by a modified HuL activity assay (Section 2.5.3) and used to calculate the mean specific productivity for each strain (Figure 35A). The method of quantification of the expression of *HAC1*, *PDI* and *KAR2* in NpαGOxZ and NpαGOxZ *Δopi1* secreting GOX was repeated for cultures of NAαT70A and NAαT70A *Δopi1* post-expression as a measure for the relative activation states of UPR in each strain while secreting HuL T70A (Figure 35B).

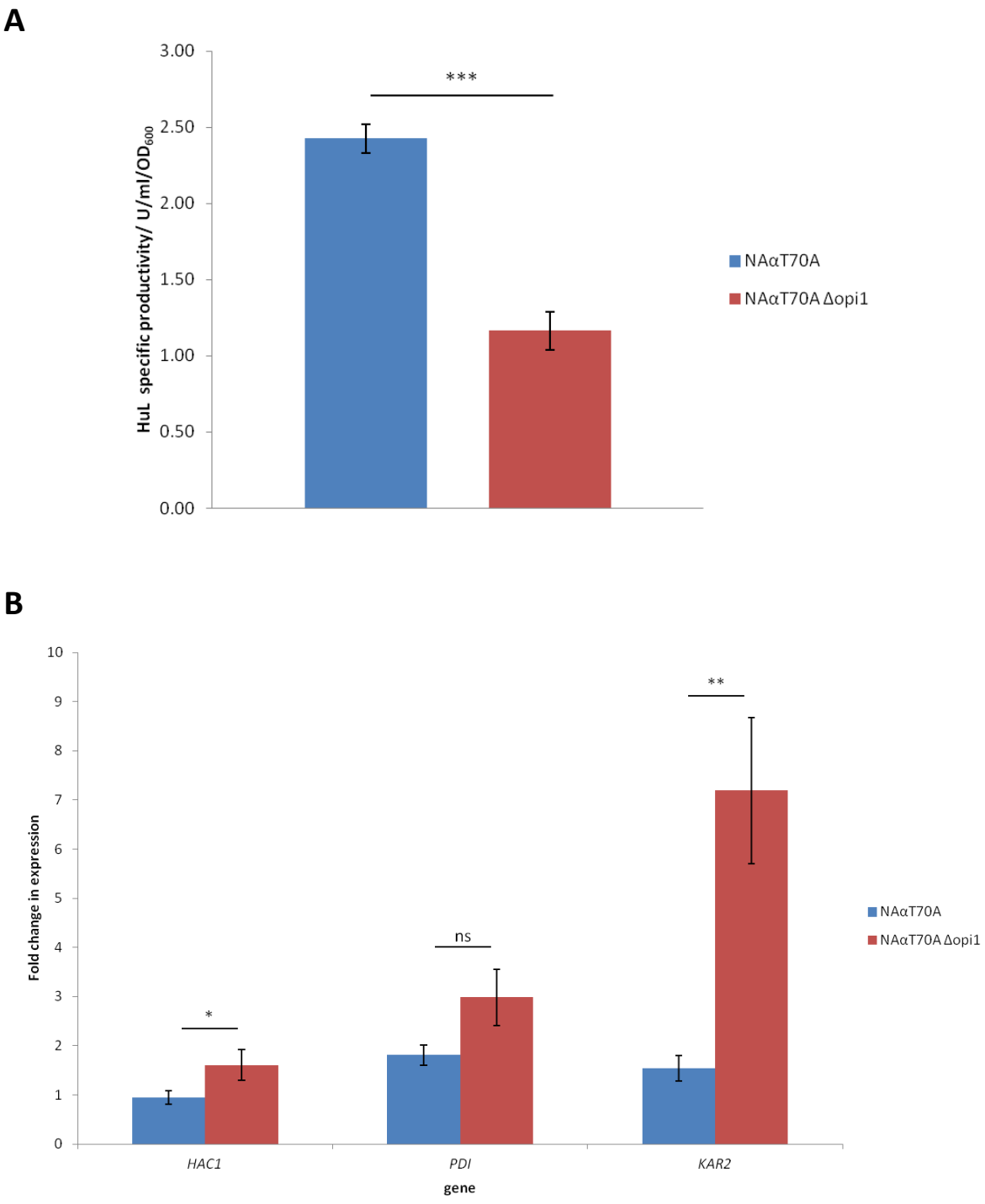


Figure 35: Mean HuL T70A titres (A) and fold change in expression of the UPR genes *HAC1*, *PDI* and *KAR2* relative to wild type NRRL 11430 levels, using *ACT1* expression as a reference (B) of NAαT70A and NAαT70A Δopi1 following a 48 hour expression trial. Error bars represent the standard error of 5 biological replicates and the significance levels of unpaired t-tests between strains for the expression of each gene are displayed – ns: $p>0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

Despite being characterised as encoding a protein inducing high levels of UPR in *P. pastoris* when targeted for cell secretion, NAαT70A expressing a single copy of *HUL T70A* did not exhibit a significant increase in the expression of *HAC1*, *PDI* or *KAR2* in comparison to the levels observed for NpαGOxZ secreting GOx. This finding was unexpected as the conditions and duration of the expression of NAαT70A was similar to those used by original study by Whyteside et al (2011). However the method of selecting each strain employed within their study involved screening a number of transformants for productivity and choosing the highest producers to continue with expression studies, without confirming the copy numbers of each transformant. Since the transformants for this experiment were selected based on the condition of containing a single copy of *HUL T70A*, it is highly likely that the clones of NAαT70A used in this experiment expressed lower levels of *HUL T70A* than in the original study, and so would not achieve the saturation of secretion that would be a likely requirement for UPR activation. The difference in specific productivity of HuL T70A and the UPR profile for NAαT70A *Δopi1* was found to mirror the results observed for the expression of aprotinin and GOx by *P. pastoris Δopi1* strains. NAαT70A *Δopi1* transformants produced approximately 50% less HuL T70A per unit of cell mass, and increased the expression of *KAR2* up to ~4.5 fold more than in native NAαT70A. *HAC1* was also significantly upregulated in NAαT70A *Δopi1* but to a lesser degree, confirming that *Δopi1* mutants experience higher degrees of UPR activation for the same level of expression of identical recombinant genes to native *P. pastoris*. The similarity in the results obtained for the production of each of the proteins tested within the study, varying in size, function and stability provide strong evidence that the deletion of *OPI1* in *P. pastoris* provides a universally adverse effect on the production of recombinant proteins and the protein processing capacity of the ER for the tested conditions.

5.5 *Δopi1* mutants exhibit altered growth characteristics on sorbitol

The lipid biosynthesis pathway in yeast, which feeds into the phospholipid biosynthesis pathway, begins with the input of acetyl-CoA derived from glycolysis or the β-oxidation of fatty acids. Acetyl – CoA is an important precursor for a range of vital cellular processes; feeding into the tricarboxylic cycle which, as well as forming part of cellular respiration, in turn provides the necessary intermediates for the biosynthesis of amino acids (Berg, Tymoczko and Stryer, 2002). Should the loss

of *OPI1* result in the constitutive derepression of phospholipid biosynthesis there could be potential ramifications for all pathways stemming from glycolysis, as an increased proportion of free acetyl-CoA is channeled towards lipid production. This provides a hypothetical explanation for the decreased productivity and heightened ER stress in *Δopi1* mutants as cells overexpressing recombinant proteins would also place a greater demand on the amino acid biosynthesis pathway, in conjunction with the hypothetically unregulated lipid biosynthesis, further limiting the availability of essential products from glycolysis in the cell. If this were the case a possible solution to alleviate the negative effects observed in *Δopi1* mutants would be to increase the availability of metabolised carbon sources during recombinant protein production. Sorbitol is a non-repressing carbon source for methanol inducible genes in *P. pastoris*, including *AOX1*, and is converted to fructose 1, 6-diphosphate which enters glycolysis. Therefore to test whether delimiting the available carbon source for glycolysis improves productivity in *P. pastoris Δopi1*, the 48 hour expression trial for NpαGOxZ and NpαGOxZ *Δopi1* described in section 5.4.2 was repeated with the indusion of 4% (w/v) sorbitol in the methanol expression medium. The final GOx titres and cell density for each culture were recorded and compared against the productivity observed in standard BMMY media (Figure 36).

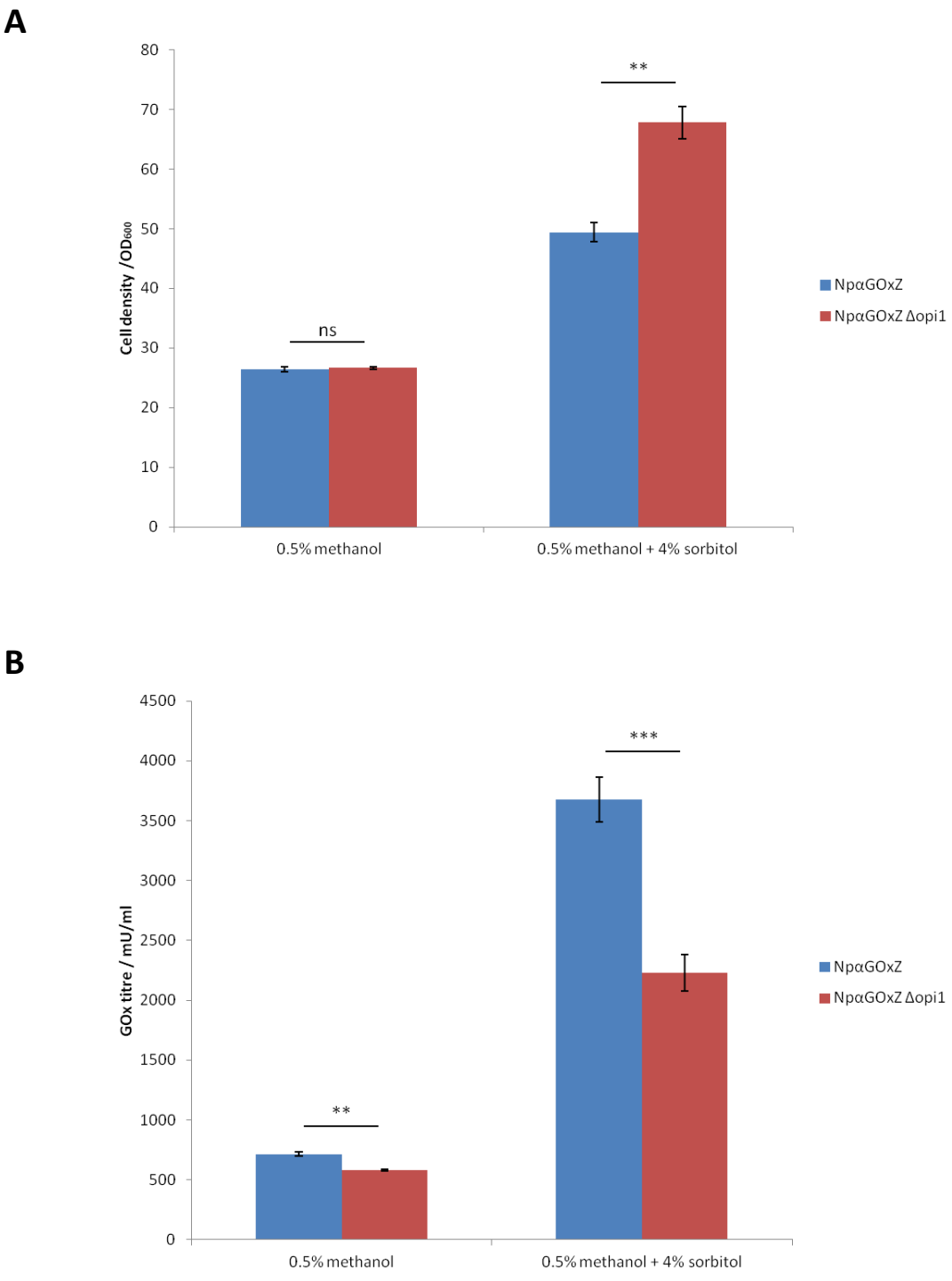


Figure 36: Final average cell density (A) and volumetric productivity (B) for cultures of *P. pastoris* NpαGOxZ and *P. pastoris* NpαGOxZ *Δopi1* following expression in BMMY (0.5% [v/v] methanol) or BMMY supplemented with 4% (w/v) sorbitol over 48 hours at 28°C. Error bars represent the standard error of 5 biological replicates and and the significance levels of unpaired t - tests between strains for the expression of each gene are displayed – ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001

Supplementing BMMY with 4% (w/v) sorbitol resulted in a large increase in the production of GOx by both NpαGOxZ and NpαGOxZ *Δopi1* due, in part, to the enhanced growth of both strains during the expression trial, although the overall magnitude of the increase was diminished in NpαGOxZ *Δopi1*. Once again NpαGOxZ *Δopi1* failed to produce comparable titres of GOx to the native strain with the addition of sorbitol, in fact producing proportionally less GOx to NpαGOxZ than when cells were grown on methanol as the sole carbon source for growth. However NpαGOxZ *Δopi1* also attained a significantly higher final cell density, producing ~1.37 fold more biomass than NpαGOxZ. Whilst the results refute the hypothesis that the poor performance of *Δopi1* mutants is attributable to an increased demand for carbon sources feeding into glycolysis, the difference observed between the final cell densities of each strain post-expression suggests a unique phenotype for sorbitol-based growth in *P. pastoris Δopi1.*

To explore the changes in the sorbitol growth characteristics of the *OPI1* deletion strain further a growth curve anlaysis was conducted between NRRL 11430 and NRRL 11430 *Δopi1* on minimal media containing 4% (w/v) sorbitol. 3 biological replicates for each strain were normalised to a starting OD_{600} of 0.1 in 25ml culture in 250ml baffled flasks. The cultures were grown at 30°C/ 250rpm, during which the cell densities for each strain were recorded at regular intervals and plotted upon the initiation of the stationary phase of growth (Figure 37).

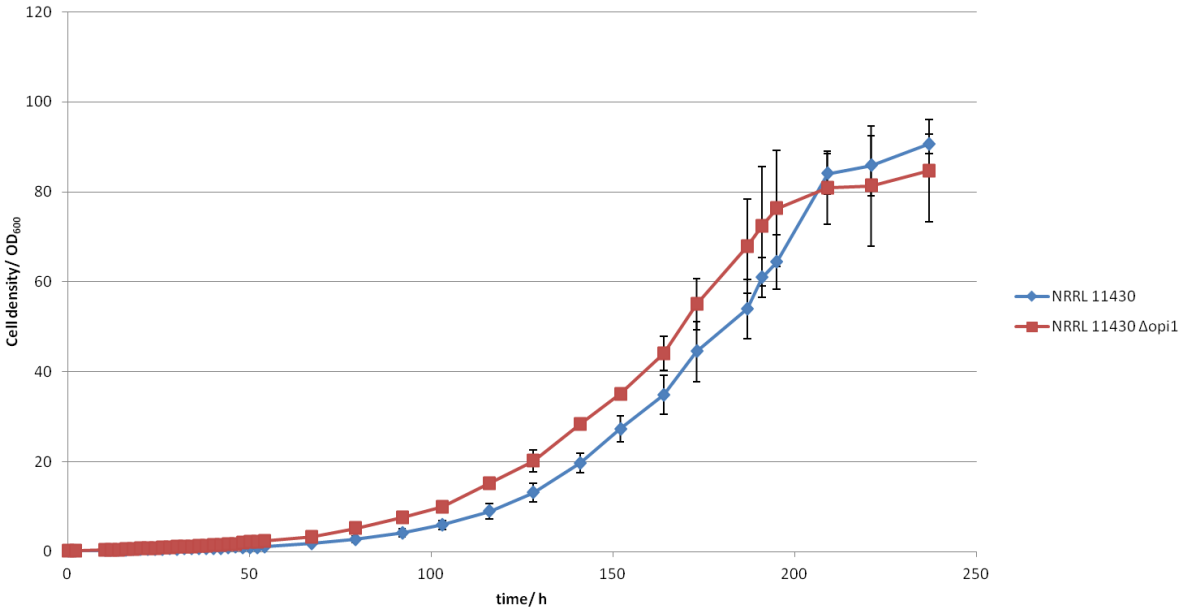


Figure 37: Growth curve of NRRL-11430 and NRRL-11430 *Δopi1* on minimal media containing 4% (w/v) sorbitol. Error bars represent the standard deviation of 3 biological replicates.

The growth rate of NRRL-11430 *Δopi1* exceeded NRRL 11430 between hours 0-80, having, achieving a mean cell density of approximately double that of NRRL 11430 after 50h.. However, once they had entered mid-exponential growth, the mean cell densities appeared to increase at equivalent rates and had equalised upon the initiation of the stationary phase in both strains. To check for any differences in the growth rate during exponential phase between NRRL-11430 *Δopi1* and NRRL 11430, the natural logarithm of the recorded absorbance values were plotted against time and the period during which the rate of growth increased linearly, represented as a straight line on the semi-log graph (9.3) (Figure A2) was identified in both strains. The starting and final absorbance values for each period were taken and used to calculate the doubling time of both NRRL-11430 *Δopi1* and NRRL 11430 with the following formula – previously described in section 4.3 and are displayed in Table 9:

$$Doubling\ time = \frac{t * \log(2)}{\log(Final\ cell\ density) - \log(Initial\ cell\ density)}$$

Table 9: Cell doubling times of NRRL 11430 and NRRL-11430 Δopi1 during growth in minimal media containing 4% (w/v) sorbitol as a carbon source.

Strain	Doubling time/ h
NRRL 11430	20.42
NRRL-11430 *Δopi1*	20.66

A ~1% difference in doubling time was found between NRRL-11430 *Δopi1* and NRRL 11430, confirming that there is only a marginal difference in the growth rates of both strains on 4% (w/v) sorbitol. One aspect that is particularly noticeable of *P. pastoris*' growth on sorbitol is that, whilst final biomass concentrations remain high, the maximum growth rate is very slow in comparison to substrates such as glucose, glycerol or methanol. This is evident from the ~20 hour doubling time of NRRL 11430 during exponential growth in addition to the duration of time that NRRL 11430 spends in lag phase growth; the semi-log graph revealing that early exponential phase is entered after 52 hours of growth. The lag phase is significantly reduced in NRRL-11430 *Δopi1*, with cells appearing to begin exponential phase growth after approximately 20 hours. Despite overall growth and growth rate remaining unchanged in *P. pastoris Δopi1*, the reduction in the lag phase before growth explains how higher cell densities were reached by NpαGOxZ *Δopi1* during a relatively short, 48 hour expression trial. However the increase in biomass during early growth in sorbitol/ methanol mixed feeds for NpαGOxZ *Δopi1* does not correlate with an increase in the productivity over the native strain indicating that *Δopi1* mutants continue to experience bottlenecks in recombinant protein secretion and increased ER stress irrespective of their improved growth characteristics on sorbitol.

5.6 Attempting to stain and visualise the ER in *P. pastoris Δopi1*

Considering the unexpected detrimental effects that the loss of OPI1 has on the productivity and the processing capacity of the ER in cells overexpressing recombinant proteins it would be pertinent to determine whether the function of Opi1p in *S. cerevisae* is conserved and, more specifically, if its

removal causes the same change in ER physiology in *P. pastoris*. To do so, a method for visualising and quantifying the surface area or volume of ER in cells would be required.

The cell permeable, lipophilic, fluorescent dye - 3,3'-dihexyloxacarbocyanine iodide ($DiOC_6(3)$) binds to lipid bilayers of mitochondria at low concentrations and, at concentrationsof 1µg/ml, has been demonstrated to stain the ER in yeast cells (Koning et al, 1993). As such, $DiOC_6(3)$ was been widely established as a standard fluorescent marker for the visualisation and study of ER morphology (Sabnis et al, 1997) and would be more suitable for the requirements of this study over more conventional methods for visualising the ER; often involving the generation of fluorescent reporters fused with ER-membrane-associated proteins. In contrast to fluorescent fusion proteins, the production of which is linked to the expression levels of their corresponding gene constructs, $DiOC_6(3)$ non-specifically binds to the ER surface, meaning that its accumulation in the cell should be directly proportional to the surface area of the ER. This relationship has been previously exploited to provide a relative quantification of changes in ER size in *S. cerevisae*, through measuring the fluorescence generated by cells stained with $DiOC_6(3)$ by flow cytometry (Block-Alper et al, 2002). The application of $DiOC_6(3)$ for the visualisation and quantification of the ER was therefore tested in *P. pastoris*.

To ascertain whether $DiOC_6(3)$ staining would provide an accurate representation of ER size in *P. pastoris*, its specificity for binding the ER was tested by visualising *P. pastoris* NRRL 11430 cells, stained with $DiOC_6(3)$, by fluorescence confocal microscopy. Cell suspensions were stained with a range of final concentrations of $DiOC_6(3)$ and imaged as described in 2.4.5 (Figure 38).

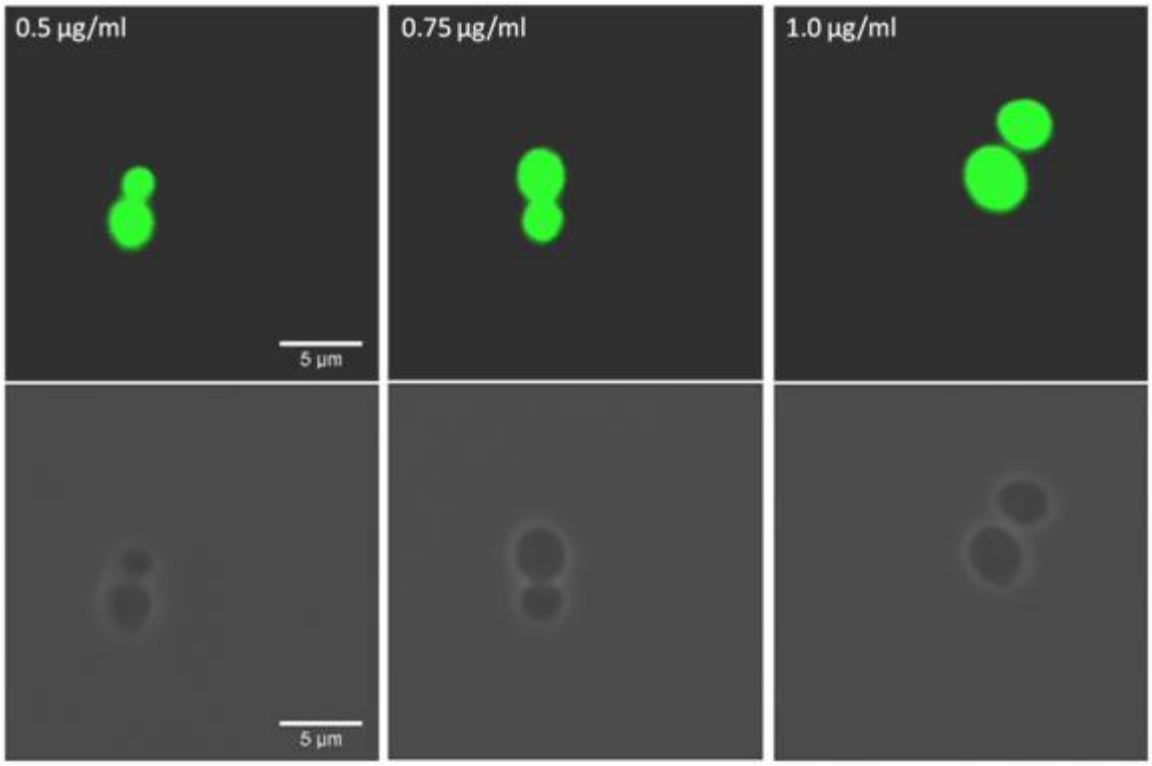


Figure 38: Fluorescence images of NRRL 11430 cells stained with a final concentration 0.5, 0.75 or 1.0µg/ml $DiOC_6(3)$ and viewed by confocal microscopy. Bright field images of the cell periphery are displayed underneath the respective fluorescence image.

Imaging of stained NRRL 11430 cells revealed that $DiOC_6(3)$ was successfully taken up and retained intracellularly. However, rather than incorporating specifically into the ER membrane, $DiOC_6(3)$ was instead found to be evenly distributed throughout the cell interior and no intracellular structures could be resolved. The dye was found to be evenly distributed through all of the focal plains examined, confirming that $DiOC_6(3)$ was intracellular as opposed to specifically staining the cell exterior, which would have been th only alternative explanation for the pattern observed in Figure 38. No differences in the pattern of staining could be observed when the final concentration of $DiOC_6(3)$ was reduced to half of the standard concentration required for ER staining in *S. cerevisae*. Whilst incubation with an excess of $DiOC_6(3)$ results in staining of the ER membrane in *S. cerevisae*, its incorporation into *P. pastoris* appears to be significantly less targeted. The fluorescence microscopy images of stained cells point towards $DiOC_6(3)$ either accumulating within the cytoplasm or non-specifically incorporating into other membrane structures including the cell membrane; the latter being more likely due to its lipophilic properties. This may provide a reason as to why $DiOC_6(3)$

staining of the ER in *P. pastoris* has not been previously reported or why no $DiOC_6(3)$ staining protocols optimised for *P. pastoris* are currently available. The lack of specificity of $DiOC_6(3)$ for the ER membrane and its high background of intracellular staining prevent its application as a marker for the estimation of ER surface area in *P. pastoris* and therefore cannot be used to quantitatively compare the difference in ER sizes between wild type and *Δopi1* cells.

5.7 Discussion

This study successfully identified and deleted a homologue belonging to the *OPI1* family of yeast transcriptional repressors in the *P. pastoris* NRRL 11430 genome. Even though cell viability was unaffected, analysis of the expression of certain UPR-associated elements showed that the *PDI* gene was significantly upregulated during regular growth in YPD media for the *Δopi1* strain. A comprehensive analysis of recombinant protein secretion involving 3 separate proteins under *AOX1*-based expression, varying in structure and function, revealed not only that the specific productivity of secreted proteins was reduced, often severely, in *Δopi1* mutants, but that the UPR genes - *HAC1* and *KAR2* were even further upregulated over native strains in the process. The regulation of *KAR2* in particular was significantly altered during recombinant protein production, as its expression was increased by a maximum of 6 fold over the native strain for the tested proteins. Considering the role of Kar2p as an ER-resident chaperone assisting in protein folding, its strong upregulation during the overproduction of secreted proteins indicates that higher quantities of misfolded proteins accumulate within the ER of *Δopi1* mutants. This is further supported by the decrease in the mean titres of extracellular proteins in *Δopi1* mutants normalised for gene copy number, suggesting that proteins that would normally traffic effectively through the secretory pathway and exit the cell in the native strain are instead obstructed within the ER to a greater degree in the *Δopi1* variant. Rather than reducing bottlenecks in the secretory pathway that occur during the overexpression of recombinant genes in *P. pastoris*, the loss of *OPI1* only seems to compound them and result in the increased activation of the UPR. Since its loss has the opposite effect to increasing the productivity of *S. cerevisae* (de Ruijter, Koskela and Frey, 2016), it is highly likely that *P. pastoris* Opi1p has evolved a different function and is directly or indirectly involved in maintaining the native processing capacity of the ER for secreted proteins. Evidence of a functional diversity of Opi1 family proteins in more

distantly related yeast species has, in fact, previously been reported. In the oleaginous yeast *Yarrowia lipolytica* the Yas3p transcription factor, identified as an Opi1p homologue, negatively regulates the transcription of genes involved in the metabolism of *n*-alkanes (Hirakawa et al, 2009) whereas in *Candida albicans* Opi1p controls the expression of both the *SAP2* protease, involved in its virulence, and genes triggering the initiation of filamentous growth (Chen et al, 2015). Additional experiments would be required to establish whether the role of *P. pastoris* Opi1p has differentiated from the *S. cerevisae* Opi1p since the method attempted to visualise the ER of *P. pastoris* in this study was unsuccessful. The fusion of fluorescent reporters to ER-resident proteins and electron microscopy remain as the 2 most viable methods for visualising the ER in *P. pastoris* (Rossanese et al, 1999; Guerfal et al, 2010). Given that *P. pastoris* Opi1p belongs to the Opi1 family and contains a leucine zipper, often involved in DNA binding, its probable role as a transcriptional regulator could also be investigated through a transcriptomics-based approach to identify any differences in global gene expression as a result of its deletion. The loss of *OPI1* has also been found to cause a dramatic change in the lipidome of *S. cerevisae*, with cells amassing higher quantities of free fatty acids and phosphatidylinositol upon the deregulation of their respective biosynthesis genes (Chumnanpuen, Nookaew and Nielsen, 2013). A similar lipidomics approach employing techniques such as gas chromatography-mass spectrometry (GC-MS) and high performance liquid chromatography (HPLC) to quantify intracellular non-polar lipids and phospholipids, could be employed in future studies to determine whether Opi1p continues to regulate lipid biosynthesis in *P. pastoris*.

Since the precise role of Opi1p in *P. pastoris* is ambiguous, though likely to not regulate ER homeostasis in the same way as *S. cerevisae* Opi1p, expanding the ER membrane through increasing the production of its constituent phospholipids remains an explorable avenue for improving the ER's processing capacity. *P. pastoris* is an inositol prototroph and contains a functional *INO1* homologue encoding the enzyme catalyzing its biosynthesis (Chi, He and Yao, 2005). Altering the expression of *INO1*, which is repressed by Opi1p in *S. cerevisae*, could therefore provide an alternative strategy for increasing the production of inositol-based phospholipids. However unlike membrane structures in *S. cerevisae*, which contain a significant inositol-derived phospholipid component (Zinser et al, 1991), the major constituents of *P. pastoris* membranes are phosphatidylcholine (PC) and phosphatidylethanolamine (PE); both making up ~27.48%/ ~36.15% of the cell membrane and ~54.4%/ 27.6% of peroxisomal membranes respectively (Wriessnegger et al, 2007; Grillitsch et al, 2014). The identification of genes involved in the synthesis of PC and PE in *P. pastoris*, or engineering

an orthologous, characterised, PC and PE biosynthetic pathway from a model organism such as *S. cerevisae* could also be considered in future approaches to achieve constitutive ER expansion.

At present, the study concludes that the deletion of *OPI1* has an adverse effect on recombinant protein secretion, reduces the processing capacity of the ER and increases ER-associated stress during recombinant gene overexpression. The function of *P. pastoris* Opi1p therefore appears to be necessary in maintaining the native protein processing and folding capabilities of the ER and so is, in contrast to Opi1p in *S. cerevisae*, beneficial to recombinant protein production.

6. Towards alleviating carbon catabolite repression of *AOX1*-based expression

6.1 Introduction

6.1.1 Aims and objectives

Due to the favourable growth characteristics of *P. pastoris* in glucose, and the number of candidate factors and pathways found to contribute towards glucose repression of P_{AOX1}, glucose repression was chosen as the primary focus of the research. The aim of this study was to ultimately generate novel strains that exhibit partial derepression of P_{AOX1} in the presence of repressive levels of glucose, and are capable of fully induced P_{AOX1}-driven expression in mixed feed media containing methanol and quantities of glucose that are detrimental to *AOX1* expression in the native host.

Firstly a simple screen differentiating between glucose repressed and methanol induced states of P_{AOX1} was designed to determine the relative production of recombinant protein from strains utilising the *AOX1* system while growing on both methanol and glucose. The screen was then implemented to test for the alleviation of glucose repression in new strains affected in 1 or more of the previously discussed factors involved in the regulation of P_{AOX1} during glucose repression and methanol induction. The strategies for alleviating glucose repression in *P. pastoris* include knockouts for genes involved in glucose repression, a number of which produce mutants already known to exhibit partial glucose derepression, or attempts to uncouple positive regulators of P_{AOX1} from their inactivation during glucose repression.

Any strains found to successfully induce *AOX1* expression in the presence of methanol and increased levels of glucose could be taken forward to expression trials for secreted recombinant proteins in a range of expression media varying in glucose: methanol ratios. The results of the study would determine whether productivity is improved over native *P. pastoris* both during standard methanol expression and for expression in media containing both glucose and methanol as growth substrates.

6.1.2 The disadvantages of methanol as both an inducing agent and growth substrate

The ability to metabolise methanol in minimal media as well as the ubiquitous use of the methanol inducible AOX1 system has resulted in the wide acceptance of standard protocols for *P. pastoris* that utilise methanol as both a carbon source for respiration and an inducer for recombinant gene expression. This aspect of *P. pastoris* expression is often viewed as advantageous as it simplifies and reduces the associated costs for producing expression media; the most common of which, such as BMMY, BMM and BMMH, include methanol as the sole carbon source. Whilst the expression of numerous proteins at commercially viable titres has been achieved in the past using this method, there is evidence to suggest that growth on methanol alone is in fact detrimental to recombinant protein production.

Firstly, the overall growth and relative growth rate of cells in methanol is poor in comparison to other substrates such as glucose and glycerol. Whilst a low growth rate presents a greater obstacle to small scale expression, in the case of *P. pastoris* it is also relevant to large scale fermentation, as studies have found a positive correlation between productivity and the specific growth rate (μ) of cells at steady state within a chemostat culture (Rebnegger et al, 2014; Stadlmayr et al, 2010). Moreover Rebnegger et al (2014) found that various genes directly involved in protein synthesis, translocation and glycosylation were upregulated with increasing μ, whereas a subset of genes comprising the proteolytic ERAD pathway were downregulated. The results implicate increasing the specific growth rate in culture as beneficial for the promotion of cellular processes related to protein production and minimising stress responses incurred by the increased rate of protein synthesis by recombinant strains. Thus the inherent concern arises that, in cultivating cells in a low growth carbon source such as methanol, a lower ceiling for the maximum specific growth rate is imposed, preventing the level of growth that would provide the optimal cell viability for protein production. Whilst direct evidence for the negative effects of growth rates in methanol would be difficult to obtain, the implementation of fermentation strategies in which methanol is co-fed with alternative carbon sources such as sorbitol and glycerol have achieved higher total growth in culture and improved productivity for recombinant β-galactosidase and heavy-chain fragment C of *botulinum* neurotoxin serotype C respectively (Zhang et al, 2003; Niu et al, 2013). The successes of both strategies at least demonstrate the merits of introducing alternative substrates promoting further growth in conjunction with methanol.

Poor growth rates are however not the only negative characteristic of growth on methanol. The requirement for oxygen as a substrate for the oxidation of methanol by AOX during the first step of the Mut pathway places a particularly high demand for dissolved oxygen during fermentation of methanol-grown *P. pastoris* at high cell densities (Chen et al, 1997; Cregg et al, 2000). During the methanol induced expression of cellulose binding module-lipase B fusion protein Jahic et al (2013) also attributed a significant loss of product with the presence of intracellular proteases in the culture supernatant, indicative of cell lysis. Similar observations were made by Sinha et al (2005) for the expression of recombinant ovine interferon-τ, in which the proteases discovered in cytoplasmic fractions and the culture supernatants were found to be vacuolar in origin, and were present at far lower concentrations in all fractions when cells were cultivated with glycerol as the sole carbon source. The results implicate an increased rate of protease production and cell lysis as a symptom of growth and expression on methanol. The introduction of oxidative species in the form of formaldehyde and hydrogen peroxide, produced from the oxidation of methanol in the Mut pathway, require immediate conversion or detoxification to preserve peroxisome integrity and prevent cell damage. Consequently the oxidative stress response in *P. pastoris* is strongly upregulated during growth on methanol, including genes encoding catalase and glutathione peroxidase (Vanz et al, 2012).

In addition to oxidative stress incurred through methanol metabolism, ER stress in response to increased protein traffic should also be taken into consideration. The transfer of cell culture to media containing methanol accompanies a substantial shift in the *P. pastoris* proteome and physiology as genes encoding proteins in the Mut pathway are activated and peroxisomes housing its primary steps are formed. Native AOX1 alone accounts for up to 30% of the total soluble protein within the cell (Couderc and Baratti, 1980) while formate dehydrogenase, S-(hydroxymethyl)-glutathione dehydrogenase (FLD1) and S-formylglutathione hydrolase (FGH1), all involved in the dissimilation of methanol into carbon dioxide, are also produced in high quantities (Vanz et al, 2012). Should total cellular protein production during methanol metabolism exceed that of growth on more preferable carbon sources, the resulting burden on the protein translation and folding machinery would be hypothetically increased even prior to the introduction of recombinant proteins. Furthermore the ER plays a primary role in the biogenesis of peroxisomes for methanol metabolism, not only in the delivery of lipid material in the formation of the organelle, but also in the targeting of certain peroxisomal proteins to localize within the peroxisome (Yan et al, 2008). Vanz et al (2012) found that production of the UPR inducible ER resident disulphide isomerase – Pdi was increased during

methanol growth, albeit for a recombinant strain overexpressing Hepatitis B surface antigen, raising ambiguity as to whether *PDI* is upregulated in response to increased ER stress due to activation of the Mut pathway or recombinant protein synthesis. However, data gathered from Edwards-Jones et al (2015) and also this project (5.3) supports the increase in transcription of *PDI* in WT *P. pastoris* during growth in rich media containing 0.5% (v/v) methanol in comparison to growth in YPD media (Figure 39).

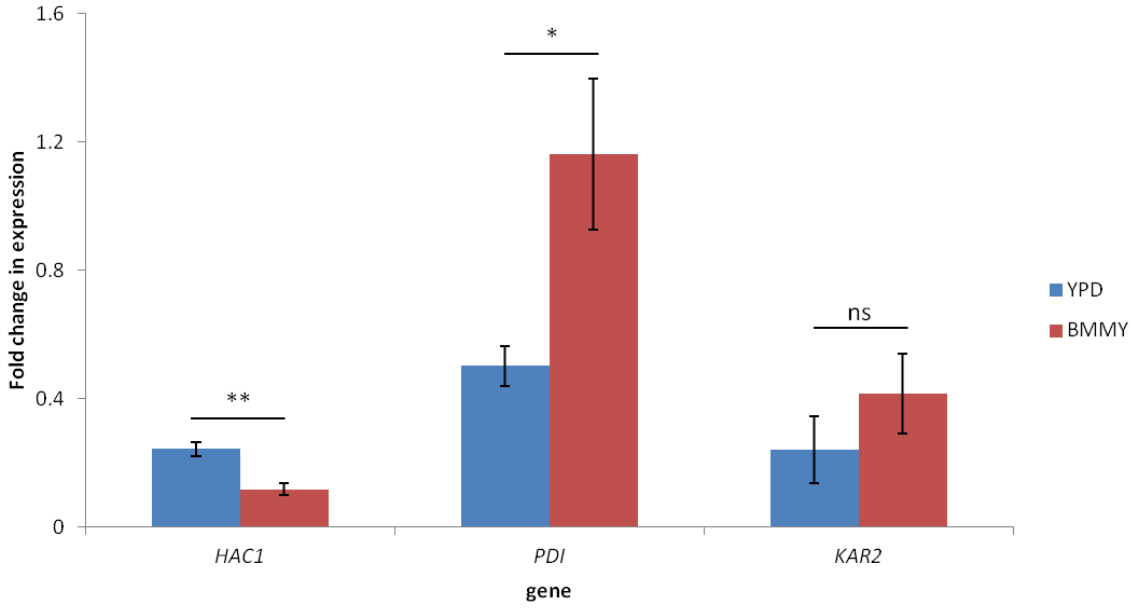


Figure 39: Expression of the UPR-regulated genes *HAC1*, *PDI* and *KAR2* by *P. pastoris* NRRL 11430 following 16 hours of growth in YPD media (2% (w/v) glucose) and BMMY media (0.5% (v/v) methanol – interpreted from 5.3. Values are given as a fold change in expression relative to *ACT1*. Error bars represent the standard error of 3 biological replicates and statistical significance values derived from unpaired Student's t-tests between each condition are reported – ns: $p>0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$.

Interestingly the expression of *HAC1* appears to be reduced in cells grown on methanol, but the overall fold change in expression is low under both conditions, although the results shown here represent total quantities of *HAC1* transcript and cannot indicate changes in the proportion of the spiced variant encoding active Hac1p. However *PDI* expression is approximately twice that of cells grown in YPD when *P. pastoris* is switched to growth on methanol. Since the role of Pdi in the ER is to catalyse the formation of disulphide bonds while reverting erroneous disulphide bonds in native

protein structures (Wilkinson and Gilbert, 2004), its increase in expression provides evidence to suggest that the protein processing capacity of the ER may be raised in response to conditions imposed by cell growth on methanol, possibly through a burden incurred by higher quantities of unfolded protein within the ER lumen. This constitutive increase in expression may be symptomatic of unfavourable conditions for recombinant protein secretion; a process which acts to only further increase the flux of proteins requiring modifications and folding into the ER, resulting in the activation of the UPR and ERAD pathways. In summary, whilst utilising methanol as the sole carbon source for cell growth and as an inducer for expression from methanol regulated promoters has been a staple within the *P. pastoris* system since its inception (Tschopp et al, 1987), research conducted in the past decade points towards a physiological background occurring exclusively during growth in methanol which is not conducive for recombinant protein production.

6.1.3 Regulation and catabolite repression of P_{AOX1}

Based on the evidence provided an ideal solution to reduce the deleterious effects of methanol-based growth would be to employ mixed feed strategies during expression, in which a different carbon source is implemented as the primary growth substrate. A more ambitious approach would involve the development of knockout strains incapable of initiating the Mut pathway, leaving methanol as an unmetabolised inducer for *AOX1* in culture in a similar vein to inducible systems in other expression systems, such as IPTG induction in *E. coli* (Kercher, Lu and Lewis, 1997). In point of fact mutants with defective methanol utilisation phenotypes – Mut^S and Mut^-, achieved through the disruption of 1 or both AOX genes, have previously been described (Cregg and Madden, 1987) and shown to increase productivity for the methanol induced expression of certain proteins (Krainer et al, 2012). However a major obstacle towards eliminating the reliance on methanol as a growth substrate, and a potential reason for the preference of Mut+ strains over Mut^S and Mut^-, is that methanol inducible promoters, particularly P_{AOX1}, are partially or fully repressed in the presence of more favourable growth substrates (Hartner and Glieder, 2006). Carbon catabolite repression (CCR) is a sophisticated, regulatory mechanism documented in both bacteria and eukaryotes and has been hypothesised to have developed as an adaptation for organisms to preferentially select more rapidly metabolisable substrates from the surrounding environment (Deutscher, 2008). In order to ensure that more readily assimilated substrates are utilised first, and to conserve the cell's metabolic machinery, genes

encoding metabolic pathways for less preferred carbon sources are often silenced during CCR. In the case of *P. pastoris* the expression of Mut pathway genes, including *AOX1*, is repressed when cells are cultured in glucose, glycerol or ethanol at concentrations that are not limiting to growth (Inan and Meagher, 2001). Non-repressing carbon sources which are metabolised by *P. pastoris* and allow for full P_{AOX1} driven expression have been identified and include sorbitol, mannitol, alanine and trehalose (Inan and Meagher, 2001). However the growth rate of *P. pastoris* in each of these substrates is considerably lower than in repressing carbon sources, limiting the number of practical growth substrates that are compatible with *AOX1* expression. In spite of their contribution towards CCR activation glucose and glycerol have previously been added in conjunction to methanol in chemostat cultures at concentrations that do not fully repress P_{AOX1}, and have succeeded in improving productivity (Paulova et al, 2012; Zhang et al, 2013). The threshold of glucose in culture required to repress P_{AOX1} has led to the development of commercially available defined growth media that maintain minimal concentrations of glucose in culture during methanol-induced expression, such as the EnPresso Y Defined growth system from BioSilta. The media includes a proprietary polysaccharide that controls the slow release of glucose in culture, promoting higher overall protein production through increased cell growth. The research findings and the presence of marketable products suggests a trend towards conducting methanol-induced expression on multiple growth substrates. However glucose and glycerol repression of P_{AOX1} continues to limit the maximum concentration of each catabolite permissible in culture to maintain *AOX1* expression, preventing the use of more optimal quantities of repressive substrate for growth in culture, as well as necessitating the requirement for accurate feed control during expression to prevent excess addition (Xie et al, 2005; Paulová et al, 2012). Engineering the partial or full alleviation of CCR of P_{AOX1} would therefore provide strains better suited to the expression of recombinant proteins on mixed feeds as they would be able to grow in media containing greater ratios of repressive substrates to methanol, promoting improved growth and viability, while retaining full P_{AOX1} activation.

The molecular interactions governing the tight regulation of the *AOX1* promoter in response to the presence of different carbon sources have yet to be understood in their entirety, although a number of studies have begun to characterise key enhancers and repressors that could provide promising candidates for the development of CCR resistant strains. Rather than being controlled by a single regulatory pathway, CCR of P_{AOX1} instead appears to comprise a complex network of distinct pathways specific to individual carbon sources, with separate modes of repression occurring during cell growth in glucose, glycerol or ethanol, as illustrated in Figure 40. Further explanation of the roles of each

factor, as well as other proteins found to be directly or indirectly involved in CCR of P_{AOX1} are summarised in Table 10.

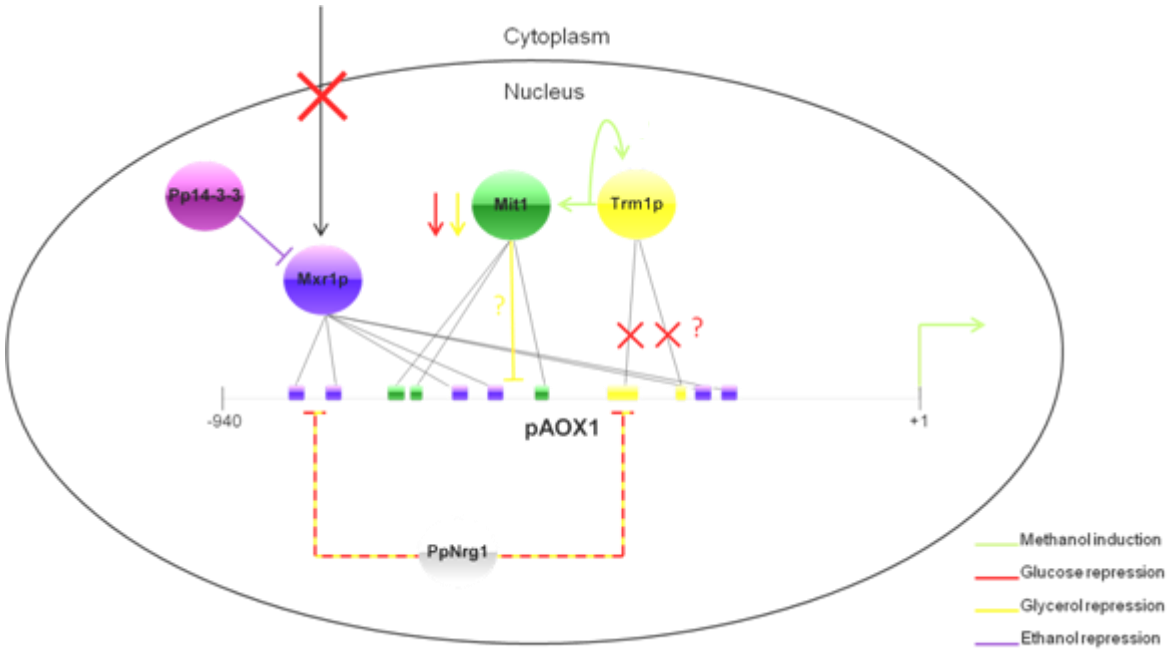


Figure 40: Schematic of the interactions between known regulatory factors for the methanol induction and CCR of P_{AOX1}. DNA binding sites for the positive transcription factors of P_{AOX1} are displayed in the correct position on the promoter and labelled in the corresponding colour of the binding factor. Coloured lines indicate interactions that occur only in response to the presence of specific carbon sources, labelled on the bottom right, while black lines indicate processes that occur irrespective of carbon source. Crosses indicate pathways that are inhibited during repression with the relevant carbon source. Question marks (?) next to interactions represent a regulatory response which has been demonstrated to occur but the precise mechanism of action is unknown.

Table 10: Factors contributing to *AOX1* regulation or CCR of *AOX1*, including a description of the phenotype of null mutants for the respective gene - if known.

	Name	Description	Mechanism of action	Mutant phenotype	Reference
Methanol induction	Mxr1p	Regulator of methanol expression	Zinc finger transcription factor essential for the activation of a number of genes involved in the Mut pathway and peroxisome formation, including AOX1. Binds directly to the AOX1 promoter.	Inability to grow on methanol or induce expression of AOX1, other Mut pathway genes and PEX genes required for peroxisome formation. Peroxisome proliferation in response to methanol cannot be observed.	Lin-Cereghino et al (2006)
Methanol induction	Trm1p	Zn(II)2Cys6-type transcription factor	Transcription factor involved in the expression of certain genes comprising the Mut pathway under methanol induction, and binds to alternative sites on the AOX1 promoter to Mxr1 or Mit1. Its binding to AOX1 is abolished in the presence of glucose. Binds to the Mit1 promoter to upregulate Mit1 expression during methanol induction. Directly binds its own promoter and also upregulates its own expression as part of a positive feedback loop.	Impaired growth in media containing only methanol as a carbon source. Reduced expression of AOX1 as well as other factors involved in the Mut pathway and peroxisome biogenesis while cultured on 0.5% (v/v) methanol.	Sahu, Krishna and Rangarajan (2014) Wang et al (2016)
Glycerol repression / methanol induction	Mit1	Zn(II)2Cys6-type methanol induced transcription factor	Transcription factor involved in the upregulation of Mut factors during methanol induction, including AOX1. Contains 2 regions characterized as redundant, named RR1 and RR3, which are required for glycerol repression of AOX1. Like Mxr1p and Trm1p, Mit1 binds to the AOX1 promoter but at independent sites to the other 2 transcription factors.	No growth on media containing methanol as the sole carbon source. 10 fold reduction in *AOX1* expression compared to wild type when grown in 0.5% (v/v) methanol. Truncation mutants missing the RR1 and RR3 domains exhibit an increase from 1.2% to 19% expression of *AOX1* compared to induced expression levels when grown in 1% (w/v) glycerol.	Wang et al (2016)

	Name	Description	Mechanism of action	Mutant phenotype	Reference
Glucose repression	PpHxt1	*Pichia pastoris* hexose transporter 1	Unknown. Hypothesised that PpHxt1 is directly involved in transcriptional repression of AOX1 during growth on glucose in conjunction to its role as a sugar carrier.	Comparable growth to wild type in glucose and methanol. Partial but significant derepression and activation of the AOX1 promoter in media containing 1% (w/v) glucose and 0.5% (v/v) methanol during early growth stages.	Zhang et al (2010)
Glucose repression	PpGss1	Glucose sensor protein homolog in *P. pastoris*	Suggested to be a low affinity glucose sensor in *P. pastoris*. Studies with 2-deoxy-d-glucose provide evidence to suggest that PpGss1 is required for glucose uptake and, consequently, glucose repression of the Mut pathway.	Severely impaired growth on 2% (w/v) glucose. Minor decrease in growth in 0.5% (v/v) methanol. Mutants capable of growing in the presence of the toxic glucose homolog - 2-deoxy-d-glucose and exhibits partial expression of AOX1 in media containing 2% glucose. Reduced glucose induced pexophagy.	Polupanov, Nazarko and Sibirny (2012)
Glucose/ glycerol repression	PpNrg1	Zn_2 Cys2His2 transcriptional repressor	Transcriptional repressor of AOX1 in the presence of glucose and glycerol. Directly binds to regions containing the 2 furthest upstream Mxr1p binding sites and 1 Trm1p binding site on the AOX1 promoter, suggesting it functions as a competitive inhibitor of the 2 transcription factors.	Reduced growth in 1% (w/v) glucose, 1% (w/v) glycerol and 0.5% (v/v) methanol. Slight de-repression of AOX1 (<25% of fully induced wild type) in media containing 0.02% (w/v) glucose and 1% (v/v) glycerol.	Wang et al (2015)

	Name	Description	Mechanism of action	Mutant phenotype	Reference
Ethanol repression	*P. pastoris* 14-3-3	14-3-3 protein homolog in *P. pastoris*	Binds to a putative 14-3-3 binding site on Mxr1p located at position 212-225 through phosphorylation of Ser215 during ethanol repression.	The point mutation S225A on Mxr1p results in the transcription factor no longer binding 14-3-3. Mutants harboring this mutation exhibit significant derepression of AOX1 in the presence of 0.5% (v/v) ethanol and comparable expression of AOX1 to the methanol induced wild type when grown in 0.5% (v/v) methanol, 0.5% (v/v) ethanol.	Parua et al (2012)
Unknown	ROP	Repressor of phosphoenolpyruvate carboxykinase	Zinc finger transcription factor originally found to be induced during methanol and biotin starvation. Localises to the nucleus and competitively binds Mxr1p recognition sites on P_{AOX1} to partially repress its expression, along with other Mut genes, when *P. pastoris* is grown in rich media containing yeast extract and peptone.	Improved growth over *P. pastoris* GS115 and increased expression of *AOX1*, *DHAS* and *FDH* when cultured in media containing 2% (v/v) methanol when supplemented with yeast extract and peptone. No difference in growth or expression during growth on minimal methanol media.	Kumar and Rangarajan (2012)
Unknown	PpZta1	Zeta crystalline homolog in *P. pastoris*	Single stranded DNA binding protein found to be bound at region -288 to -115 on the AOX1 promoter, which is inhibited in the presence of NADPH. No other information on its expression profile in different carbon sources or its mechanism of action in relation to catabolite repression is available, although its ability to bind P_{AOX1} alone suggests it is involved in AOX1 regulation.	Unknown.	Kranthi, Balasubramanian and Rangarajan (2006)

Of each of the 3 known modes of CCR of P_{AOX1} – glucose, glycerol and ethanol repression, the largest number of contributing factors and pathways have been elucidated for glucose repression at present. The expression profile of *AOX1* consists of a catabolite repressed state, an induced state caused by the presence of methanol and absence of repressive carbon sources, and a derepressed state under the absence of both methanol and repressive carbon sources, through which expression occurs at approximately 1-2% of the induced state (Tschopp et al, 1987). Methanol induction is facilitated by the collective interaction of 3 zinc finger transcription factors Mxr1p, Trm1p and Mit1 with P_{AOX1}, each of which are regulated differently to achieve the characteristic tight control of *AOX1* expression in response to carbon source availability.

6.1.3.1 Mxr1p

Mxr1p was the first global transcriptional regulator for Mut pathway genes discovered for *P. pastoris*, and is essential for their activation during growth on methanol (Lin-Cereghino, 2006). Its association with *AOX1* regulation was initially perceived during screening experiments for peroxisome biogenesis mutants incapable of growth in methanol, during which *mxr1* mutants defective in both growth in methanol and *AOX1* expression were also identified (Lin-Cereghino, 2006). Not only is AOX1 production severely reduced to less than 5% of native levels in methanol induced *Δmxr1* strains, but the production of the Mut pathway enzymes Fdh and Fld, as well as the production of catalase for hydrogen peroxide detoxification, are also negatively affected (Lin-Cereghino, 2006).

Mxr1p is a large, 1155 amino acid protein with homology to the *S. cerevisiae* alcohol dehydrogenase regulator - Adr1p and, like Adr1p, contains a conserved, zinc finger DNA binding domain located within its N-terminal region, specifically at positions 38-95 (Lin-Cereghino, 2006). Through sequential yeast one-hybrid assays a major activation domain essential for transcription was mapped to positions 246-280 in Mxr1p (Parua et al, 2012). The same study also found that a truncated form of Mxr1p consisting of the N-terminal 400 amino acids, spanning both the zinc-finger DNA binding domain and the major activation domain was sufficient to generate native Mxr1p activity in *mxr1* mutants, despite the positioning of a conserved fungal transcription factor regulatory region, predicted by NCBI's conserved domain database (Marchler-Bauer et al, 2015), beyond of the specified amino acid range (Figure 41).

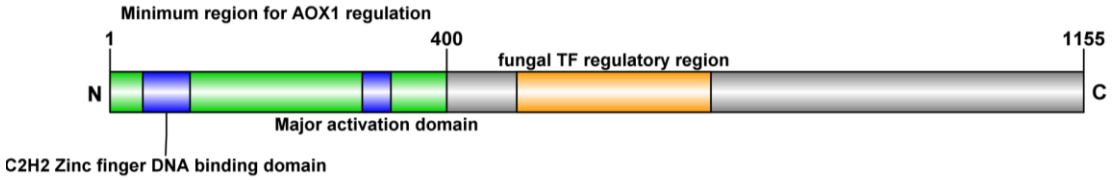


Figure 41: Linear map of Mxr1p, containing the positions of defined, essential regions for function (blue), predicted domains (orange) and the minimal region sufficient for *AOX1* regulation (green) derived experimentally.

This finding is of particular importance for informing genetic engineering approaches entailing the recombinant expression or alteration of *MXR1* to modify *AOX1* expression, as it describes a more minimal, simplistic form of Mxr1p with potentially fewer regulatory domains that can be expressed to achieve *AOX1* activation as well as reducing the size of the associated biological parts and expression vectors that would be required.

Electrophoretic mobility shift assays (EMSA) between the N-terminal 150 amino acids of Mxr1p, expressed recombinantly in *E. coli*, and DNA fragments spanning the P_{AOX1} region identified 6 specific Mxr1p binding sites on P_{AOX1} (Kranthi et al, 2009). A core binding motif for Mxr1p of 3'-CYCC-5' was determined through sequence analysis of the binding domains, and confirmed when mutations within the motifs abolished Mxr1p binding. The study notes that the presence of multiple Mxr1p binding sites on P_{AOX1} is likely to be a key factor involved in the exceptionally high strength of P_{AOX1}, particularly as deletions of regions containing a fraction of the total number of Mxr1p binding sites on the promoter significantly reduces *AOX1* expression *in vivo* (Hartner et al, 2008).

The regulation of Mxr1p to reduce its binding and prevent its activation of P_{AOX1} in glucose – grown cells appears to occur through the alteration of its subcellular localisation. During growth of *P. pastoris* on glucose Mxr1p is found diffused throughout the cell interior whereas, during growth on other tested carbon sources including methanol, glycerol and ethanol, it appears exclusively within

the nucleus where it can activate transcription (Lin-Cereghino, 2006). The established importance of its DNA binding for *AOX1* expression, and the fact that it remains localised to the nucleus in the presence of the other repressive carbon sources – ethanol and glycerol, suggests that the change in Mxr1p's subcellular localisation in glucose-grown cells, and its subsequent reduction in nuclear concentration, is one of the key underlying mechanisms that is unique to glucose repression of P_{AOX1}.

6.1.3.2 Trm1p

The methylotrophic yeast *Candida boidiinii* contains an analogous Mut pathway to *P. pastoris* that is also regulated by methanol induction and CCR (Hartner and Glieder, 2006). A putative gene found in *P. pastoris* encoding a protein with 58% homology to Trm1p, a transcriptional activator of at least 5 Mut pathway genes in *C. boidiinii* (Sasano et al, 2008), was studied and also named Trm1p – although it is sometimes referred to Prm1p in reference to the organism (Sahu, Krishna and Rangarajan, 2014). *Δtrm1* mutants exhibited retarded growth in methanol, albeit not as severe as *Δmxr1* mutants, reduced expression of the Mut pathway genes AOX1, FLD, DHAS, FDH and reduced expression of 3 peroxisome biogenesis genes, implicating Trm1p as a second master regulator of methanol induced expression in *P. pastoris* (Sahu, Krishna and Rangarajan, 2014). An EMSA of Trm1p did not identify any Trm1p binding sites on P_{AOX1}, although EMSAs performed in a separate study found that Trm1p bound P_{AOX1} at 2 independent sites from Mxr1p (Wang et al, 2016). The EMSAs were complemented with *in vivo* chromatin immunoprecipitation (ChIP) analysis of Trm1p binding to P_{AOX1} in response to different carbon sources, in which it was found that Trm1p binding was abolished in glucose – grown cells, suggesting a potential regulatory mechanism governing Trm1p activity during glucose repression. The deletion of *TRM1* resulted in an 80% drop in transcription of the gene encoding the 3rd positive regulator of P_{AOX1} – *MIT1* implying that Trm1p can indirectly influence AOX1 activation through upregulating expression of *MIT1* (Wang et al, 2016). This was supported by the identification of a Trm1p binding site on the *MIT1* promoter (Wang et al, 2016). Expression of recombinant GFP from the *TRM1* promoter was also reduced in *Δtrm1* mutants grown in glucose, glycerol and methanol, and Trm1p was also found to bind a site on its own promoter (Wang et al, 2016). The results of the study indicate that Trm1p enhances *AOX1* expression during methanol induction via 3 interactions: binding P_{AOX1} to increase *AOX1* transcription, upregulating *MIT1* expression and upregulating its own expression. The latter pathways imply that, through the activity of Trm1p, the

production of both Mit1 and Trm1p increase concurrently in a positive feedback loop during methanol induction.

6.1.3.3 Mit1

Mit1 is the most recently characterised of the 3 known zinc finger transcriptional activators of *AOX1*, having been identified as a potential candidate for Mut pathway regulation in *P. pastoris* through the sequence homology of its DNA binding domain to the Mut pathway transcriptional regulator Mpp1 in *Hanselula polymorpha* (Prielhofer et al, 2015; Wang et al, 2016). Its null mutant also exhibits the same characteristic poor growth on methanol and highly downregulated *AOX1* expression as *Δmxr1* and *Δtrm1* mutants (Wang et al, 2016). 3 specific Mit1 binding sites were located on P_{AOX1} at independent locations from those of both Mxr1p and Trm1p, suggesting that all 3 transcription factors independently activate *AOX1* transcription to varying degrees (Wang et al, 2016). Interestingly a number of domains on Mit1, named RR1, RR3 and UR3 were found to be involved in glycerol repression of P_{AOX1}, as their deletion resulted in significant derepression of *AOX1* expression in glycerol-grown cells, expressing *AOX1* at up to 19% of methanol induced levels (Wang et al, 2016). The study concluded that the presence of these sites confirm that Mit1 is involved in the direct repression of P_{AOX1} during growth on glycerol but it is also possible that the domains house recognition sites for *trans* – acting factors that repress *AOX1* expression by binding and inactivating Mit1, similar to ethanol repression of Mxr1p.

While the expression of both MXR1 and TRM1 genes are maintained at consistent, low levels regardless of carbon source availability (Lin-Cereghino, 2006; Sahu, Krishna and Rangarajan, 2014), meaning that they are regulated post – translationally during CCR, the regulation of Mit1 occurs at the transcriptional level. Mit1 remains permanently localised to the nucleus and is capable of binding P_{AOX1} in both glucose and glycerol-grown cells, but transcription of the *MIT1* gene is heavily repressed in both growth substrates (Wang et al, 2016). Upon shifting cells to growth on methanol, transcript levels of *MIT1* increase sharply, most likely due in part to the positive feedback loop in expression caused by Trm1p, suggesting that repressed and induced states exist for *MIT1* expression (Wang et al, 2016).

6.1.3.4 Negative regulators

In contrast to the number and level of characterisation of positive enhancers of P_{AOX1} discovered less is known about *trans* – acting factors that directly repress P_{AOX1} during CCR. A homologue for ScNrg1, a transcription factor mediating glucose repression in *S. cerevisiae*, was found in *P. pastoris*, and encodes a zinc finger transcription factor (Wang et al, 2015). The protein, named PpNrg1, was found through EMSA to bind P_{AOX1} at sites between the Mxr1p binding sites furthest upstream of *AOX1* and a site directly spanning 1 of the 2 Trm1p binding sites (Wang et al, 2015). *Δppnrg* mutants also displayed increased *AOX1* expression on 1% (v/v) glycerol and 0.02% (w/v) glucose to up to approximately 17% of methanol induced levels. An increase in expression was also observed for a number of other Mut pathway genes including *AOX2*, *DAS*, peroxisome biogenesis genes and 2 genes encoding peroxisomal membrane proteins, supporting the role of PpNrg1 as a global transcriptional repressor of the Mut pathway during glycerol and glucose repression (Wang, 2015). However it should be noted that the repressive effects of PpNrg1 were not observed in cells grown in 1% (v/v) glucose, indicating that its function is more likely to be specific to glycerol repression and under growth limiting conditions. The location of the PpNrg1 binding sites on P_{AOX1} indicate that PpNrg1 must compete for binding with the positive transcription factors Mxr1p and Trm1p as its mechanism of repression during CCR.

During ethanol repression Mxr1p was found to be inactivated post – translationally by a 14-3-3 homologue in *P. pastoris* (Parua et al, 2012). The interaction was initially investigated based on evidence that the homologue of Mxr1p – Adr1p is regulated by the 14-3-3 protein Bmh in *S. cerevisiae*, which was shown to also be capable of binding Mxr1p (Parua et al, 2012). Sequence alignments of both Adr1p and Mxr1p revealed the homologous positioning of a serine residue (S225) on Mxr1p that is phosphorylated on Adr1p during 14-3-3 interaction. The point mutation S225A resulted in an Mxr1p variant that no longer bound a 14-3-3 homologue in *P. pastoris*, and strains bearing the mutation exhibited significant levels of *AOX1* transcription in media containing ethanol as a sole carbon source and an almost complete alleviation of repression in media containing both ethanol and methanol. Interestingly a ChIP-seq analysis revealed that Mxr1p is still able to bind P_{AOX1} during ethanol repression, suggesting that the *P. pastoris* 14-3-3 is capable of non – competitively inhibiting Mxr1p activity after it has bound to recognition sequences (Parua et al, 2012). The study

was able to confirm a second inhibitory mechanism for Mxr1p which provides an explanation, in part, as to how CCR continues to occur in spite of conditions where Mxr1p is fully localised in the nucleus.

Other proteins found to affect glucose repression of P_{AOX1} include the low affinity hexose transporter PpHxt1 and the low affinity glucose sensor PpGss1 as deletion mutants for either gene were capable of expressing *AOX1* at minimal levels when cells were grown in media containing glucose, although growth in glucose was also impaired (Zhang et al, 2010; Polupanov, Nazarko and Sibirny, 2012). The phenotype of their respective knockout mutants, and their function in *P. pastoris* suggests that both PpHxt1 and PpGss1 indirectly facilitate glucose repression by their involvement in maintaining intracellular levels of glucose; a factor which most likely signals the activation of glucose repression pathways.

6.2 Design of a screen to determine P_{AOX1} activation states

In order to confirm the alleviation of glucose repression in the novel strains generated within the experiment a suitable screen allowing for the partial quantification of the expression of recombinant genes by P_{AOX1} in response to glucose and methanol is required. The expression of the zeocin resistance marker *sh ble* provided a suitable candidate to screen for P_{AOX1} activity since the inactivation of zeocin within cells is titratable to the quantity of Sh ble produced. The resistance of *P. pastoris* cells expressing *sh ble* under P_{AOX1} to increasing concentrations of zeocin would therefore correlate with the level of transcriptional repression experienced by P_{AOX1}. For the generation of test strains expressing sh ble under P_{AOX1} the vector pAVECRS, based on the *P. pastoris* expression vector pAVE522, was designed and assembled (Figure 42).

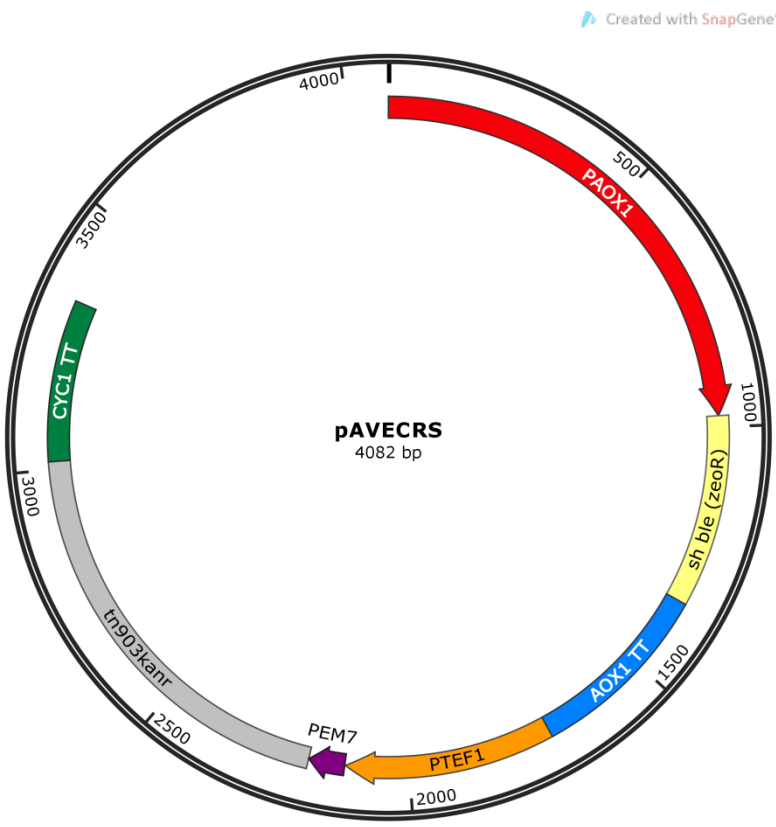


Figure 42: Vector map of pAVECRS, for the selection of strains with alleviated repression of P_{AOX1} based on zeocin resistance.

6.2.1 Cloning

pAVE522 was selected to form the vector backbone for pAVECRS as it contains the *AOX1* promoter followed immediately by a multiple cloning site to allow the simple construction of *AOX1* expression cassettes. However pAVE522 also contains *sh ble* under the constitutive *TEF1* yeast promoter as a selectable marker, requiring replacement with an alternative marker for the vector to function correctly as part of the screen. *sh ble*, including some flanking sequence belonging to P_{TEF1} and the *CYC1* terminator region was excised from pAVE522 by a double restriction digest with NcoI and StuI. The $G418^R$ marker tn903kanr was PCR amplified from the $G418^R$ cassette supplied by Fujifilm Diosynth Biotechnologies (Billingham, UK), with ends containing the regions of PTEF1 and the CYC1 TT removed by the NcoI and StuI digest, as well as a further 40bp of complementary Gibson ends, using 10-Gibkan-F and 11-Gibkan-R (Figure 43).

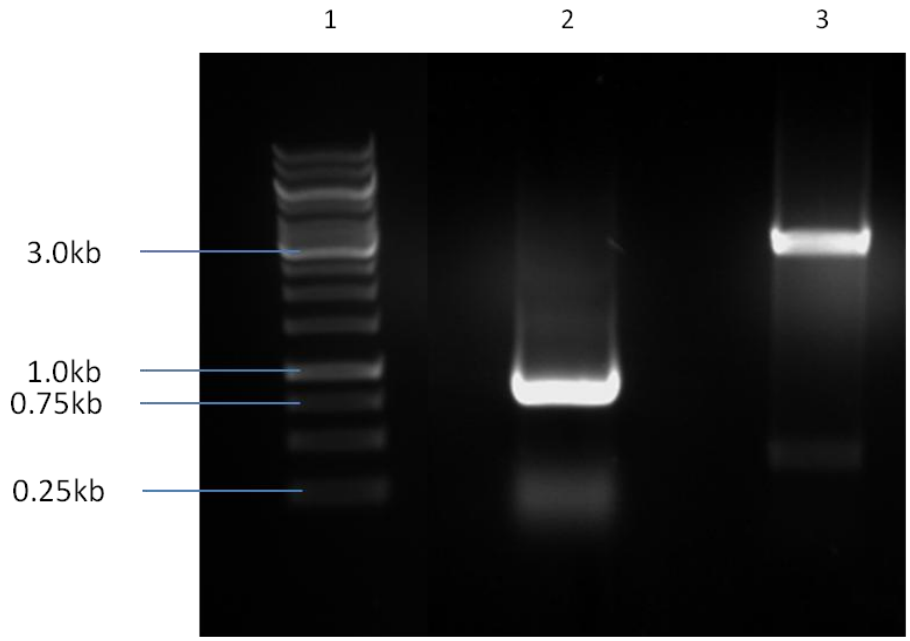


Figure 43: Generation of parts to assemble the vector backbone for pAVECRS. Lane 1 – 1.0kb DNA marker, 2 – PCR of tn903kanr with replacement sequence and complementary ends to PTEF1 and CYC1 TT, 3 – removal of a ~400bp fragment encoding sh ble and flanking sequence from pAVE522 by restriction digest with NcoI and StuI.

tn903kanr was inserted into the site formerly containing *sh ble* and the new vector, named pAVE522-k, was electroporated into *E. coli* with selection on 30µg/ml kanamycin. The correct insertion of tn903kanr downstream of PTEF1, and the restoration of the 3' region of P_{TEF1} were confirmed by sequencing with the primers 16-Selseqb-F and 17-Selseqb-R.

Following its assembly, pAVE522-k was restricted with BstBI and SacII to remove the *S. cerevisiae* α-mating secretion signal and to linearise the vector between the *AOX1* promoter and the *AOX1* TT. *sh ble* was then PCR amplified using the Gibson primers 24-GibAOXzeo-F and 25-GibAOXzeo-R from pAVE522, gel purified and inserted downstream of PAOX1 in the linearised pAVE522-k by Gibson assembly. Colony PCRs of *E. coli* cloned with the assembled vector and sequencing within the AOX1 locus with 12-5'AOX1 and 13-3'AOX1 confirmed the correct insertion of *sh ble* with high efficiency to complete pAVECRS (Figure 44).

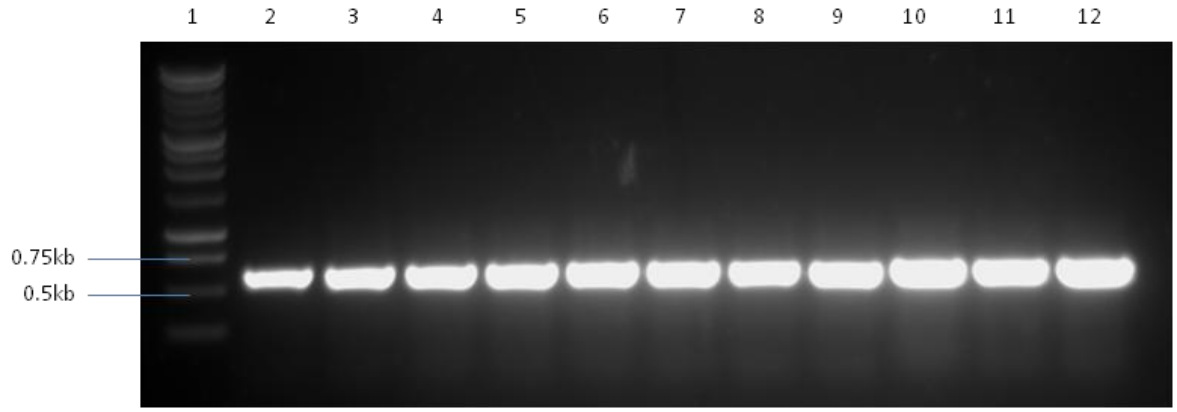


Figure 44: Colony PCRs confirming the insertion of sh ble downstream of P_{AOX1} in pAVECRS for each tested *E. coli* colony. Lane 1 - 1.0kb DNA marker, 2-12 colony PCRs of individual *E. coli* + pAVECRS colonies with 12-5'AOX1 and 13-3'AOX1

6.2.2 Resistance of *P. pastoris* cloned with pAVECRS to zeocin is dependent on P_{AOX1} derepression and induction

To generate a screenable strain for P_{AOX1} activity, pAVECRS was linearised by SacI and integrated into the *AOX1* locus of *P. pastoris* NRRL 11430 by electroporation with selection on G418 sulphate. To test whether the expression of *sh ble* could be influenced by glucose repression of P_{AOX1} a positive *P. pastoris* NRRL 11430 + PAVECRS transformant was grown in 5ml minimal media containing either 1% (w/v) glucose or 1% (w/v) glucose and 0.5% (v/v) methanol and grown for a maximum period of 16 hours to prevent the concentration of glucose in culture falling to growth – limiting levels. The transformants were also grown separately in 1% (w/v) sorbitol to examine the derepressed state of P_{AOX1} and 0.5% (v/v) methanol to fully induce *sh ble* expression. The cultures were normalised to a low OD_{600} of 0.1 to prevent tolerance arising from the sequestering of free zeocin in the media by an excess of cells, and were subsequently plated onto 0-500µg/ml zeocin gradient agar plates matching the growth substrate used in each culture. The growth on each carbon substrate was imaged after 48 hours and is displayed in Figure 45.

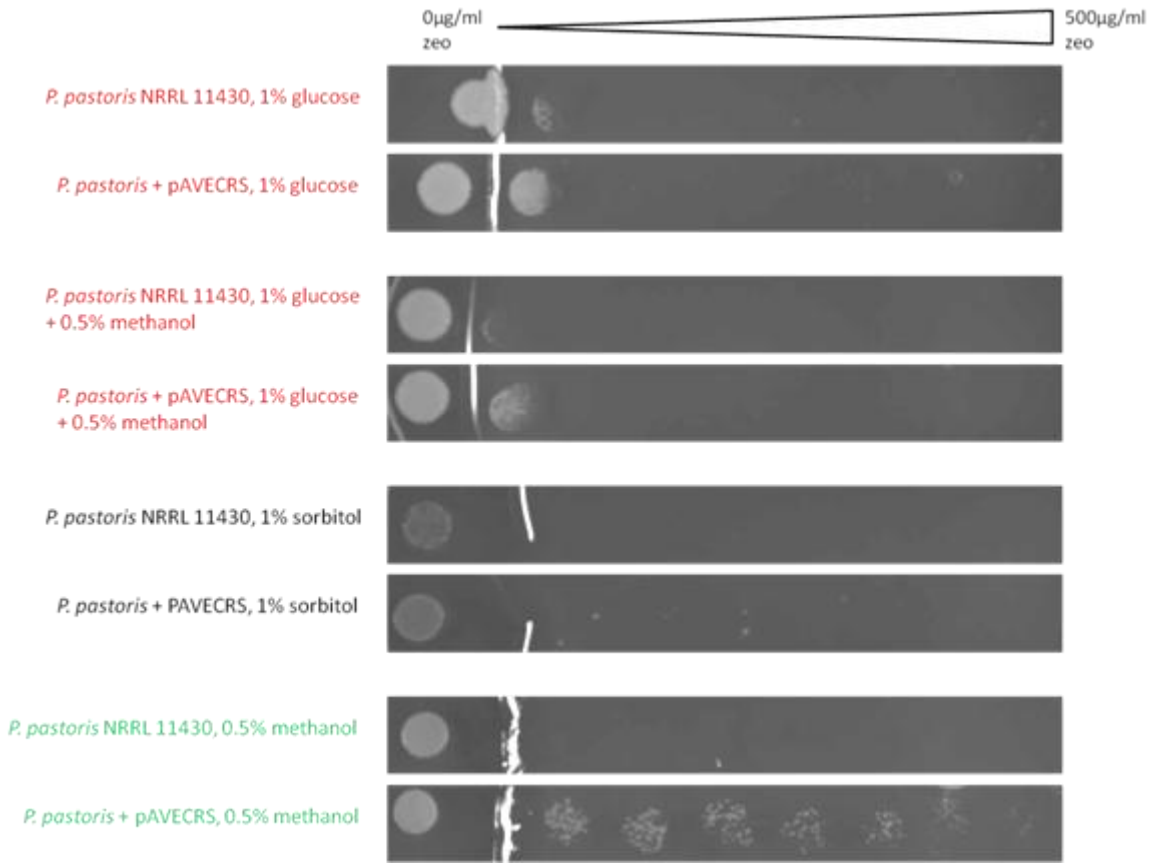


Figure 45: Susceptibility of *P. pastoris* + pAVECRS to zeocin grown on different carbon sources. Media containing 1% glucose (red) represses the *AOX1* promoter even in mixed feeds containing methanol. Insufficient Sh ble is expressed under de-repressing conditions during growth on sorbitol (black) and inducing conditions in which methanol is the only carbon source present is sufficient to generate partial resistance to up to 500µg/ml zeocin (green).

The addition of an excess of glucose in the media, even in the presence of methanol, was sufficient to prevent resistance to zeocin developing in *P. pastoris* expressing *sh ble* under P_{AOX1}, as *P. pastoris* + pAVECRS grew at only a slightly higher concentration of zeocin than the wild type in the presence of glucose . The finding suggests that a low, basal level of expression occurs even under glucose repression. Despite the lack of growth substrates inducing CCR for *P. pastoris* + pAVECRS cultured in 1% (w/v) sorbitol no resistance to even low concentrations of zeocin could be observed, suggesting that expression levels from derepressed PAOX1 do not generate adequate quantities of active Sh ble to confer resistance to zeocin. A decrease in the susceptibility to zeocin compared to *P. pastoris* NRRL 11430 could only be observed in *P. pastoris* + pAVECRS under methanol induction, which resulted in zeocin resistant colonies developing at concentrations ranging up to 500µg/ml. However colony

growth and morphology was relatively impaired in comparison to cells grown in the control region of plates containing no zeocin. While resistance was evidently increased, the comparison indicates that only a fraction of the cells plated grew to form colonies. A contributing factor to the non-uniformity of resistance between cells was attributed to the low growth rates of *P. pastoris* in minimal methanol media containing which, when coupled with exposure to high concentrations of zeocin, results in a substantial decrease in viability and weakened growth even in resistant cells. The addition of 3% (w/v) sorbitol to the minimal methanol media significantly improved the confluency of methanol induced, resistant *P. pastoris* + pAVECRS cells on zeocin gradient plates, and was therefore used for the remainder of the zeocin gradient plate experiments in this study.

The ability to distinguish methanol induction from repression of P_{AOX1} by examining the growth of *P. pastoris* + pAVECRS on zeocin gradient agar plates has established a suitable, simple screen for the alleviation of glucose repression from *AOX1* – based expression. The presence of 1% (w/v) glucose in the growth media is able to silence the expression of *sh ble* in wild type *P. pastoris* transformed with pAVECRS and therefore novel strains exhibiting significant alleviation in glucose repression of P_{AOX1} should be identifiable through their increased resistance to zeocin when grown in mixed feeds of glucose and methanol.

6.3 Attempting to implement a CRISPR system for the disruption of multiple genes in *P. pastoris*

To date the majority of research into CCR of *P. pastoris* has entailed the deletion or mutation of single genes with hypothetical roles in CCR and examining its effect on the regulation of methanol inducible genes. The discovery of glucose transporters, sensors, transcriptional repressors of P_{AOX1} and regulatory regions on Mxr1p that can be deleted or altered to partially alleviate CCR of P_{AOX1} suggests that a combination of a number of the modifications previously studied would yield further alleviation of glucose repression over the respective single mutant strains. However, in an attempt to find the optimal set of mutations to uncouple a regulatory network as complex as glucose repression from *AOX1* expression, a combinatorial approach testing multiple gene deletions within single strains would be required, necessitating the use of a more sophisticated system for gene disruption or editing than has previously been used within this project. Such a system would have to employ

scarless gene modification to preserve the limited number of antibiotic selectable markers available to clone expression vectors into the novel *P. pastoris* strains generated. A method requiring fewer integration/ self-removal events of gene specific knockout cassettes, typical of FLP-recombinase gene deletion systems, would also facilitate more rapid generation of multiple deletion strains, providing an improvement to current systems validated for *P. pastoris*. Therefore, to design a system for the generation of multiple candidate strains for the alleviation of glucose repression, and to contribute to the molecular toolbox of *P. pastoris*, the development of a rudimentary CRISPR-Cas9 system for the disruption of multiple genes was explored.

6.3.1 CRISPR – Cas9 in yeast

Clustered, regularly interspaced short palindromic repeats (CRISPR) coupled with CRISPR – associated factors (Cas) were originally discovered as an adaptive immune response in different species of bacteria and archaea for excising invasive viral DNA from the host genome (Barrangou et al, 2007), but have since been developed into a powerful *in vivo* gene editing tool. The modularity of the CRISPR-Cas system, its high targeting efficiency for specific DNA sequences within host genomes and recent modifications expanding its repertoire to include systems for inserting foreign DNA into loci, as well as the transcriptional activation or silencing of target genes, has generated considerable interest within the biotechnology community (Larson et al, 2013; Zhang et al, 2015). Currently CRISPR-Cas systems have been developed for a host of model organisms including the yeasts *S. cerevisiae* and *Candida albicans* (DiCarlo et al, 2013; Sander and Joung, 2014; Vyas, Barrasa and Fink, 2015).

The most common use for CRISPR-Cas is the disruption of genes within the host genome to prevent their expression, with the CRISPR-Cas9 system, adapted from the native type II CRISPR system in *Streptomyces pyogenes*, having emerged as a standard for this application (Jinek et al, 2012; Cong et al, 2013; Mali et al, 2013). CRISPR-Cas9 is made up of 2 elements – the first being the *Streptococcus pyogenes CAS9*, which encodes a large DNA endonuclease. The endonuclease activity of Cas9 is significantly reduced unless paired with a guide RNA (gRNA), which consists of a "scaffold" sequence necessary for Cas9 binding and a targeting region, designed by the user, to direct the Cas9-gRNA complex to the target DNA sequence and initiate double stranded breaks (Cong et al, 2013; Mali et al, 2013) (Figure 46).

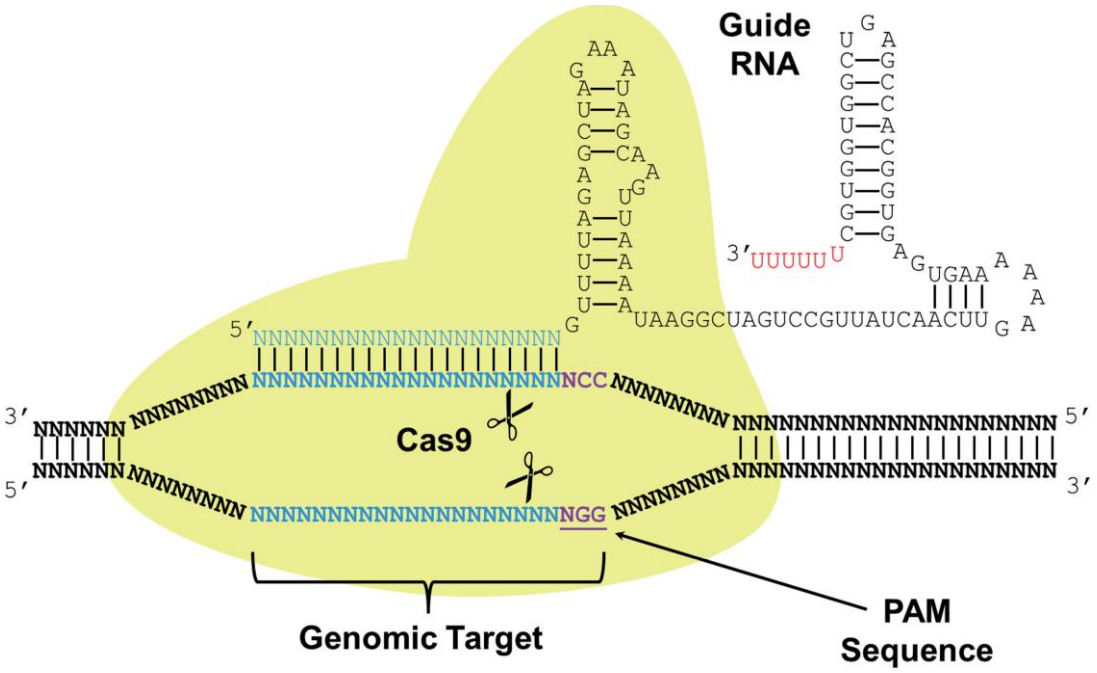


Figure 46: Schematic of gRNA interacting with Cas9 to direct site – specific double stranded DNA breaks in the target sequence, taken from DiCarlo et al (2013).

The gRNA is modified to target the site specific activity of Cas9 by altering the targeting region to complement a 20bp sequence located within the gene to be disrupted (Jinek et al, 2012). The one limitation that exists for the targeting of Cas9-gRNA complexes within the genome is that the 20bp recognition sequences occur directly 5' to an NGG triplet, known as the protospacer adjacent motif (PAM) (Mojica et al, 2009; Jinek et al, 2012). Once Cas9 has been successfully targeted to the desired region, mutations within the region causing a frameshift or a premature translation termination of the encoded gene, occurring as a result of the repairing of double stranded DNA breaks by the error prone non-homologous end joining pathway (NHEJ) can be screened for in the affected cells (Cong et al, 2013; Mali et al, 2013). Alternatively precise editing of the region spanning the recognition and break site can be achieved by co-transforming the CRISPR-Cas9 machinery into cells with homologous donor DNA (Ran et al, 2013). The donor DNA encodes the precise sequence modification required by the user, such as a premature stop codon, flanked by homologous sequence to the target gene adjacent to the CRISPR-Cas9 recognition site. The donor DNA then forms a template to stimulate the repairing of the Cas9-mediated DNA damage by the cell's homology direct repair (HDR) pathway, which results in the insertion of the modification contained in the donor DNA, into the genome with high fidelity (Ran et al, 2013).

The relatively small size of the gRNA and the target region means that different gRNA expression cassettes targeting Cas9 to multiple regions on the genome are simple and cost effective to construct. The stable expression of gRNA in cells is not necessarily a requirement either, as genome editing by transforming cells with the genes encoding Cas9 and gRNA on transient expression cassettes that are not replicated within the host has been validated in a number of model organisms (Jiang et al, 2014) (Rahdar et al, 2015; Zhang et al, 2016). This strategy has also been partially implemented in *S. cerevisiae*, where the targeted mutation of 2 individual genes at high efficiency was achieved in cells stably expressing *CAS9* under the constitutive *TEF1* promoter, following transformation with transient gRNA expression cassettes and donor templates (DiCarlo et al, 2013).

The prospect of inducing the disruption of specific genes through the transient expression of gRNAs in *P. pastoris* would be ideal for the rapid generation of large numbers of multiple knockout strains. The transient CRISPR-Cas9 system, if successfully implemented in *P. pastoris* with high targeting efficiencies, would enable the sequential or even multiplex targeting of several genes by transforming CAS9-expressing cells with different combinations of gRNAs. As such this study will attempt to construct a semi-transient CRISPR-Cas9 system, comparable to the system previously established in *S. cerevisiae*, for multiple gene disruption in *P. pastoris*.

6.3.2 Constitutive expression of *CAS9* compatible with pPICZ/ pAVE522 vectors

To begin developing a CRISPR – Cas9 system in *P. pastoris*, a strain stably expressing *CAS9* under a strong, constitutive promoter was required. Since any mutants generated through CRISPR-Cas9 editing would be subsequently transformed with pPICZ or pAVE522 – based vectors, including the glucose repression screening vector pAVECRS, the use of a *CAS9* expression vector integrating into the *AOX1* locus, utilising zeocin or G418 resistance as selectable markers would impede the further testing of the resultant strains for the alleviation of glucose repression. Considering these requirements, a CAS9 expression vector, based on the vector pGrzαHSA (Aw, 2012), was designed (Figure 47).

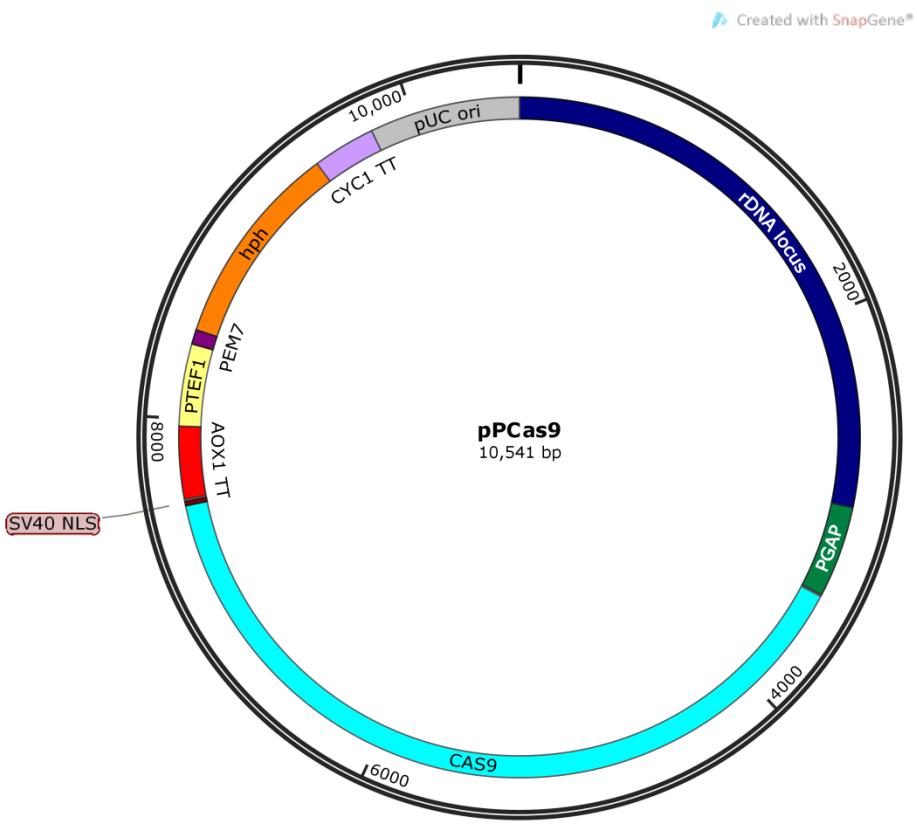


Figure 47: Vector map for pPCas9, containing the human optimised *CAS9* with C-terminal SV40 nuclear localisation sequence downstream of the *GAP* promoter for constitutive expression in *P. pastoris*.

pGrzαHSA integrates within the ribosomal DNA (rDNA) locus and contains a the gene encoding human serum albumin with the *S. cerevisiae* α – mating secretion signal, expressed under the constitutive *GAP* promoter. The replacement of *αmf – HSA* with the *CAS9* gene tagged with a C-terminal SV40 nuclear localisation signal (NLS), confirmed to effectively target recombinant proteins to the nucleus in *P. pastoris* (Weninger, Glieder and Vogl, 2015), and the replacement of sh ble with *hph*, encoding the hygromycin resistance gene from *Klebsiella pneumonia* would create a *CAS9* expression vector compatible with pPICZ and pAVE522 – based vectors.

6.3.2.1 Cloning

A double restriction digest of pGrzαHSA with LguI and NotI removed *αmf – HSA* downstream of P_{GAP}, in addition to ~200bp of the 3' region of P_{GAP}. This region was included as a Gibson part in order to restore P_{GAP} upon reassembly of the vector with CAS9, through PCR amplification of the P_{GAP} 3' region with the primers 50-GibGapfrag-F and 51-GibGapfrag-R (9.1). Finally human optimised *CAS9-SV40* was PCR amplified from p414-TEF1p-Cas9-CYC1t, with flanking MssI restriction sites to enable the use of the vector backbone for future studies and complementary ends to its locus of insertion within the vector using the primers 52-GibCas9-F and 53-GibCas9-R (9.1) (Figure 48).

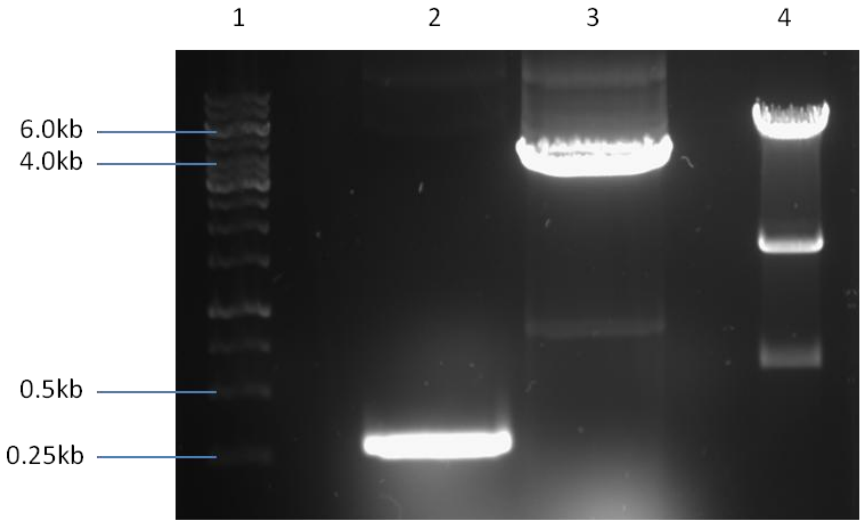


Figure 48: Generation of parts for the assembly of *CAS9* into pGrzαHSA. Lane 1 - 1.0kb DNA marker, 2 – PCR amplicons of ~200bp of 3' PGAP with Gibson ends, 3 - PCR amplicons of human optimised CAS9-SV40 with MSSI restriction sites and Gibson ends, 4 – Restriction digest of pGrzαHSA with LguI and NotI with the ~6.0kb band corresponding to the linearised vector backbone.

The 3 parts were joined by Gibson assembly and transformed into *E. coli* BioBlue with selection on 25µg/ml zeocin. The insertion of CAS9-SV40 and the restoration of P_{GAP} was confirmed by sequencing the plasmids isolated from a subset of positive transformants with the primers 111-5'GAP and 13-3'AOX1.

To replace *sh ble* with *hph*, pGrzαHSA containing *CAS9-SV40* was restricted with the enzymes NcoI and StuI, and *hph* was PCR amplified with complementary ends to the former *sh ble* site on pGrzαHSA with 48-Gibhph-F and 49-Gibhph-R. Gibson assembly was used again to insert *hph* into the vector, which was then cloned into *E. coli* BioBlue with selection on 100µg/ml hygromycin – b. Sequencing the plasmids isolated from the resulting positive transformants with 16-Selseqpb-F and 17-Selseqb-R confirmed the correct insertion of *hph* to complete the *CAS9* expression vector, which was named pPCas9.

pPCas9 was linearised within the rDNA sequence with SpeI, transformed into *P. pastoris* NRRL 11430 by electroporation, and transformants were selected on 200µg/ml hygromycin b. To confirm the constitutive expression of *CAS9-SV40* from the rDNA locus, which had not previously been tested within this study, total cellular RNA was purified from *P. pastoris* +pPCas9 grown for ~16 hours in 5ml YPD. An RT-PCR of the isolated RNA with the primers 55-Cas95'1kb-F and 56-Cas95'1kb-R, annealing specifically to the 5' ~1kb region of *CAS9*, confirmed the presence of the CAS9-SV40 mRNA transcript, indicating its expression from P_{GAP} (Figure 49).

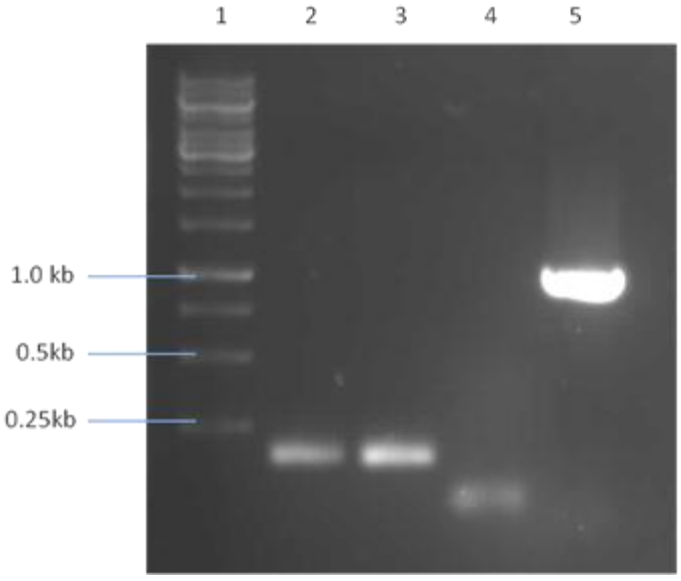


Figure 49: RT-PCR of *P. pastoris* NRRL-11430 cloned with pPCas9 to confirm transcription of *CAS9*. Lane 1 – 1kb DNA size marker, 2 – control of *P. pastoris* NRRL-11430 cDNA with primers amplifying ~200bp of *ACT1*, 3 – *P. pastoris* NRRL-11430 + pPCas9 cDNA with primers for *ACT1*, 4 – negative control of *P. pastoris* NRRL-11430 cDNA with primers amplifying 1kb of the 5' region of *CAS9*, 5 - *P. pastoris* NRRL-11430 + pPCas9 cDNA with primers amplifying 1kb of the 5' region of *CAS9*.

6.3.3 Design and transformation of gRNA expression cassettes targeting *ADE2*

With the establishment of a strain constitutively expressing Cas9, assumed to be nuclear localised by the SV40 signal sequence confirmed to function in *P. pastoris* by Weninger, Glieder and Vogl (2015), expression cassettes for the transcription of mature, functional gRNA were designed for the targeting of Cas9 to the *P. pastoris ADE2* homologue. *ADE2* encodes a phosphoribosylaminoimidazole carboxylase in *S. cerevisiae* and *P. pastoris* that catalyzes an essential step during purine biosynthesis (Stotz and Linder, 1990) (Lin Cereghino et al, 2001). In media containing low levels of adenine, *ade2* mutants produce a visible phenotypic change, as the build-up of coloured, purine precursors within the vacuoles of cells results in the growth of red/pink colonies on solid media (Lin Cereghino et al, 2001). *ADE2* therefore provides an ideal candidate gene to test the functionality and targeting efficiency of the CRISPR – Cas9 system as a proof of concept.

The web tool E-Crisp (http://www.e-crisp.org/E-CRISP/) was used to locate all possible 20bp Cas9 – gRNA recognition sequences with either a 3' NGG motif or 3'NAG, which has also been demonstrated to function as a PAM, in *P. pastoris ADE2*. The recognition sequences were aligned against the *P. pastoris* genome to control for potential off target Cas9 activity, leading to the identification of 2 unique recognition sites on *ADE2* occurring closest to the translation start site: a 5'-N_{20}NAG-3' at position 340 and a 5'-N_{20}NGG-3'at position 689 (9.4.1).

The production of functional gRNAs in eukaryotes requires the use of promoters that do not recruit RNA polymerase II (RNA Pol II) to drive expression. The promoters for protein encoding genes all direct RNA Pol II – mediated transcription of their respective genes, which processes the resulting RNA post-transcriptionally to form mature mRNA for translation (Kornberg, 1999). These post – transcriptional modifications, including the attachment of the 5' cap and 3' poly-adenylation, interfere with gRNA function and so CRISPR – Cas systems in eukaryotes often employ promoters recruiting RNA Pol III, which transcribes non – coding RNAs such as tRNAs and components of the spliceosome natively (Ma et al, 2014). This presents an obstade to the expression of gRNAs in *P. pastoris*, as there are no well characterised or tested RNA Pol III promoters for recombinant expression in the *P. pastoris* molecular toolkit. To circumvent this issue 2 separate strategies were devised for the design of *P. pastoris* gRNA expression cassettes.

6.3.3.1 S. cerevisiae P_{SNR52} and SUP4 TT

SNR52 encodes a small, nucleolar non – coding RNA in *S. cerevisiae* and its native promoter has been previously utilised to drive RNA Pol III mediated transcription (Guffanti et al, 2006). P_{SNR52} was also used for the expression of functional gRNAs in the *S. cerevisiae* CRISPR – Cas9 system, which the current system tested within this study is modelled on (DiCarlo et al, 2013). Based on the compatibility observed for a number of native *S. cerevisiae* promoters with gene expression in *P. pastoris* as a rationale, the *S. cerevisiae* gRNA expression cassette, comprising P_{SNR52} and the *SUP4* RNA Pol III specific transcription terminator was tested for gRNA expression within the *P. pastoris* CRISPR – Cas9 system. Whilst P_{SNR52} has yet to be tested, the SUP4 TT has been shown to retain its function when implemented in *P. pastoris* (Young et al, 2009).

6.3.3.2 Ribozyme – mediated processing for mature gRNA production

A second strategy, designed initially by Gao and Zhao (2014), facilitates the expression of gRNAs from RNA Pol II promoters. DNA encoding the gRNA is flanked by sequences encoding a Hammerhead (HH) type ribozyme at the 5' end, and the hepatitis delta virus (HDV) ribozyme at the 3' end. Following transcription, the HH and HDV ribozymal units within the mature mRNA molecule form secondary structures that catalyze their self – cleavage, removing the 5' and 3' ends from the mRNA molecule and releasing the full gRNA free of any post – transcriptional modifications. As such a gRNA cassette was designed with HH and HDV ribozymes, using the G1 promoter, a variant of P_{GAP} (Qin et al, 2011) to drive high level constitutive expression, and the *S. cerevisiae ADH1* transcription terminator (9.4.3) (Figure A4).

Expression cassettes employing each strategy, for separate gRNAs targeting either of the 2 previously identified recognition sites on *ADE2* were designed and synthesised in *E. coli* holding vectors. The total 4 expression cassettes were designated as shown in Table 11. Further information about the composition of each cassette can be found in 9.4.

Table 11: List of the 4 different gRNA expression cassettes for ADE2 disruption utilising either the 340-5'N20-NAG-3' or 689-5'N20-NGG-3' recognition sequences, and either the *S. cerevisiae* or ribozyme flanking strategies for gRNA expression.

ADE2 target	*S. cerevisiae* CRISPR – gRNA cassette	HH and HDV ribozyme cassette
340 - NAG	P_{SNR52}-340-SUP4t	P_{G1}-340-ADHt
689 - NGG	P_{SNR52}-689-SUP4t	P_{G1}-689-ADHt

6.3.3.3 Transformation of gRNA cassettes

The transient gRNA cassettes were transformed into *P. pastoris* + pPCas9 either within circular *E. coli* holding vectors, or as linear PCR products amplified with 57-sc-gRNA-F/ 58-sc-gRNA-R for the *S. cerevisiae*-based gRNA cassettes or 59-ribo-gRNA-F and 60-ribo-gRNA-R for cassettes utilising the HH and HDV ribozymes. After transformation cells were incubated at 28°C in 5ml YPD for 12 hours to allow the expression of both *CAS9* and gRNAs. The incubation period was not extended further, and the cell samples were minimally agitated at 100rpm to prevent excessive growth as *ade2* variants are slower growing, and risk being outcompeted by unmutated cells. 10-fold serial dilutions of cell suspensions post-incubation up to a dilution factor of 10^6 were set up in PBS, and plated in quadruplicate on YPD + 200µg/ml hygromycin-b to yield ~500 discrete colonies per transformation. The various conditions tested for each gRNA cassette are summarised in Table 12.

Table 12: List of the conditions, including the subset of gRNAs tested for each condition, for the transformation and expression of transient gRNA cassettes in *P. pastoris* + pPCas9.

Cloning method	Expression cassette format	Donor template	gRNAs tested
Electroporation (Thermo Fisher Scientific)	Circular plasmid/ linear PCR product	No	P_{SNR52}-340-SUP4t P_{SNR52}-689-SUP4t P_{G1}-340-ADHt P_{G1}-689-ADHt
Electroporation (Fujifilm)	Circular plasmid/ linear PCR product	No	P_{SNR52}-340-SUP4t P_{SNR52}-689-SUP4t P_{G1}-340-ADHt P_{G1}-689-ADHt
Electroporation (Fujifilm)	Linear PCR product	Yes	P_{SNR52}-689-SUP4t P_{G1}-689-ADHt
Electroporation (Lithium acetate + DTT pretreatment)	Linear PCR product	No	P_{SNR52}-340-SUP4t P_{SNR52}-689-SUP4t P_{G1}-340-ADHt P_{G1}-689-ADHt
Electroporation (Lithium acetate + DTT pretreatment)	Linear PCR product	Yes	P_{SNR52}-689-SUP4t P_{G1}-689-ADHt

To test whether precise gene editing through the HDR pathway would be possible as well as simple gene disruption through the error-prone NHEJ pathway, a ~120bp donor template was designed to contain a premature stop codon in-frame, within the NGG PAM located at position 689 on *ADE2*, and synthesised for use in conjunction with the P_{SNR52}-689-SUP4t and P_{G1}-689-ADHt gRNA cassettes. Under conditions requiring its inclusion, 1nmol of the double stranded donor template was co-transformed into *P. pastoris* + pPCas9 with 5-10µg of the appropriate gRNA.

None of the transformation conditions yielded any colonies displaying the *ade2* phenotype with any of the 4 gRNA cassettes tested, or with the inclusion of a donor template. To determine whether the transformation efficiency for the insertion of the gRNA cassettes was limiting their expression within cells, a variety of electroporation methods were also applied. Both the standard protocol specified by Thermo Fisher Scientific for the transformation of *P. pastoris* and a modified electroporation protocol from Fujifilm Diosynth Biotechnologies (Billingham, UK) were tested. The electrocompetency of *P. pastoris* + pPCas9 was also improved by incubating competent cell suspensions in 0.1M lithium acetate and 10mM DTT prior to electroporation, which was observed to increase the transformation efficiency of *P. pastoris* by up to 150 – fold over conventional protocols (Wu and Letchworth, 2004). However none of the electroporation protocols used resulted in the generation of any colonies with the ade2 phenotype, suggesting that transformation efficiency alone was not an impediment to the expression of CRISPR-Cas9. To check for the presence of any synonymous mutations or mutations not abolishing Ade1p function, the ADE1 locus was PCR amplified from the genomic DNA of 10 colonies cloned through the lithium acetate + DTT pretreatment method with either P_{SNR52}-689-SUP4t or P_{G1}-689-ADHt, with or without the donor template. Sequencing of *ADE2* with 63-ade2seq-F and 64-ade2seq-R revealed that the *ADE2* gene from colonies for each of the specified conditions remained unchanged from the wild type, indicating the failure of the formation of the Cas9-gRNA complex *in vivo*, or its targeting of *ADE2*.

The remaining possibilities explaining the lack of CRISPR-Cas9 activity against *ADE2* could include that transient expression is not viable for the production of adequate quantities of gRNAs for effective Cas9 targeting, in which case the use of a CRISPR-Cas9 system would be limited for the requirements of future experiments within this study. Another possibility could be that the Cas9 protein is incorrectly folding within the *P. pastoris* environment, or that translation rates are lowered due to the human optimised version of *CAS9* introducing combinations of rare codons for *P. pastoris*. Whilst low levels of activity of the CRISPR-Cas9 system in some cells could potentially be occurring, the results of the experiment are conclusive in that the actual gene disruption efficiency of the proof of concept CRISPR-Cas9 system is far too low (<0.02%) to use in practice for the knockout of genes involved in glucose repression, all of which would produce no visible phenotype and thus would require more intensive screening. As such, an alternative method for the generation of glucose repression mutants would have to be developed.

6.4 Partial alleviation of glucose repression by the constitutive expression of *MIT1*

6.4.1 Strategies for uncoupling positive transcription factors of *AOX1* from glucose regulation

In the absence of an effective gene knockout system for the removal of known factors contributing towards the transcriptional repression of P_{AOX1}, the focus of the study was shifted to the transcriptional enhancers of *AOX1* expression during methanol induction. Previous strategies looking to mitigate the inhibition of the positive *AOX1* transcription factors during glucose repression have achieved success through constitutively expressing *MXR1* and *TRM1* under the *GAP* promoter (Takagi et al, 2012). Strains overexpressing both *MXR1* and *TRM1* can partially induce *AOX1* in the presence of non-limiting quantities of glucose. Since Mxr1p is distributed evenly throughout the cell interior during glucose repression, increasing its production most likely provides more similar levels of nudear Mxr1p present during methanol induction to competitively bind P_{AOX1}. Increasing the expression of *TRM1* constitutively also partially uncouples its carbon source-dependant regulation as Trm1p upregulates its own transcription, most likely as a mechanism to activate *AOX1* expression – but only during methanol induction (Wang et al, 2016). Currently the strategy of constitutively expressing *AOX1* positive transcription factors to reduce glucose repression has not been tested on *MIT1*, though it arguably bears more significance to *MIT1* than it does for either *MXR1* or *TRM1*. The regulation of Mit1 differs from Mxr1p as it is localised to the nudeus during all modes of CCR and, unlike Trm1p, is also capable of binding P_{AOX1} in glucose-grown cells (Wang et al, 2016). Since only its expression is affected during glucose repression, it could be concluded that Mit1 regulation occurs purely at the transcriptional level. Therefore constitutively expressing *MIT1* should result in fully uncoupling its activity from glucose repression, rather than simply compensating for its inhibition. In methanol-grown cells Trm1p also actively upregulates *MIT1* transcription (Wang et al, 2016), so constitutively expressing *MIT1* would also have the benefit of emulating a Trm1p function during methanol induction. To investigate the effect of constitutively expressing MIT1 on glucose repression of P_{AOX1}, the following vectors were designed:

6.4.1.1 Constitutive MIT1 expression – pPhGmit1

To generate a strain constitutively expressing *MIT1* with compatibility to pAVECRS to test for glucose repression alleviation, and pPICZ/pAVE522 expression vectors, a MIT1 expression vector integrating outside of the AOX1 locus and utilising selectable markers other than zeo^r and G418^r was required. As such the vector backbone created for pPCas9, integrating into the rDNA locus and using hygromycin resistance as a selectable marker, was used to design the *MIT1* expression vector – named pPhGmit1 (Figure 50).

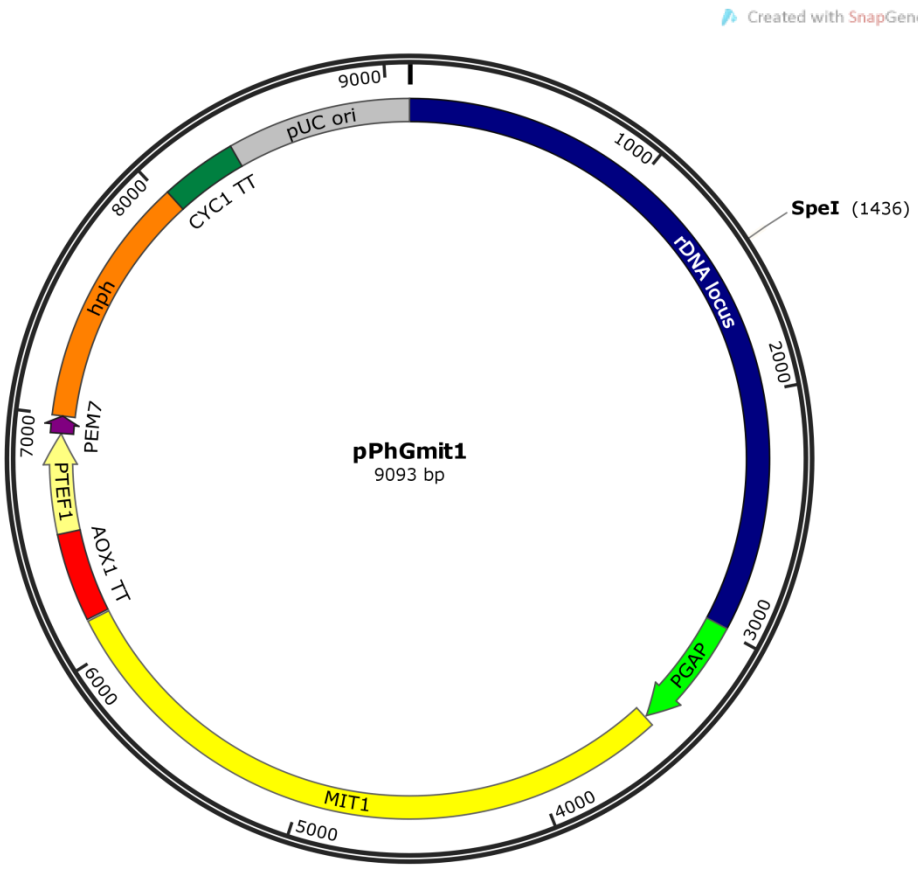


Figure 50: Vector map of pPhGmit1 for the constitutive expression of *MIT1* with compatibility to pPICZ/pAVE522 expression vectors.

6.4.1.2 Constitutive MIT1 and MXR1 expression – pPhGmm1

The deletion of the gene encoding Mxr1p impairs *AOX1* expression and cannot be rescued through overexpressing *MIT1*, suggesting that both transcription factors induce P_{AOX1} independently, and that the uncoupling of both Mit1 and Mxr1p from glucose inhibition will be necessary to alleviate the repression of P_{AOX1} (Wang et al, 2016). Since the mechanism by which Mxr1p changes its subcellular localisation in response to glucose is unknown, the overexpression of *MXR1* under P_{GAP} was opted for as a strategy to overcome low nuclear concentrations of Mxr1p during glucose repression. However the limited number of remaining available expression vectors and selectable markers necessitates expressing both *MXR1* and *MIT1* in a single, polycistronic vector as opposed to two separate vectors. The inclusion of both of the commonly used strong, constitutive promoters for *P. pastoris* – P_{GAP} and P_{TEF} currently in pPhGmit1 presents a second design constraint as repeating either promoter in the vector to express *MXR1* would introduce closely located homologous sequences, which could compromise the vector through loop-out recombination once integrated into the host genome. As a solution the viral T2A sequence was implemented in the design of a polycistronic vector for the co-expression of both *MIT1* and the sequence encoding the functional region of Mxr1p (Figure 51). Viral 2A sequences were discovered originally in picornavirus polyproteins (Ryan, King and Thomas, 1991) as short, ~20aa peptides that cause the host ribosome to skip the synthesis of a bond at their C-termini during translation and resume for any sequence immediately downstream, resulting in the production of multiple, discrete peptides from a single mRNA transcript (Radcliffe and Mitrophanous, 2004). 2A sequences have been established as tools for polycistronic vectors for over a decade (Szymczak et al, 2004) and those specifically tested in *P. pastoris* include the T2A peptide from the *Thosea asigna* virus, which was used to engineer a biosynthetic pathway comprising of 9 genes within a single expression cassette (Geier et al, 2015). The T2A sequence was therefore selected to be cloned immediately downstream of *MIT1*, followed by the 5' 1200bp of *MXR1*, encoding a minimal functioning version of Mxr1p, in frame to facilitate their simultaneous expression by P_{GAP}.

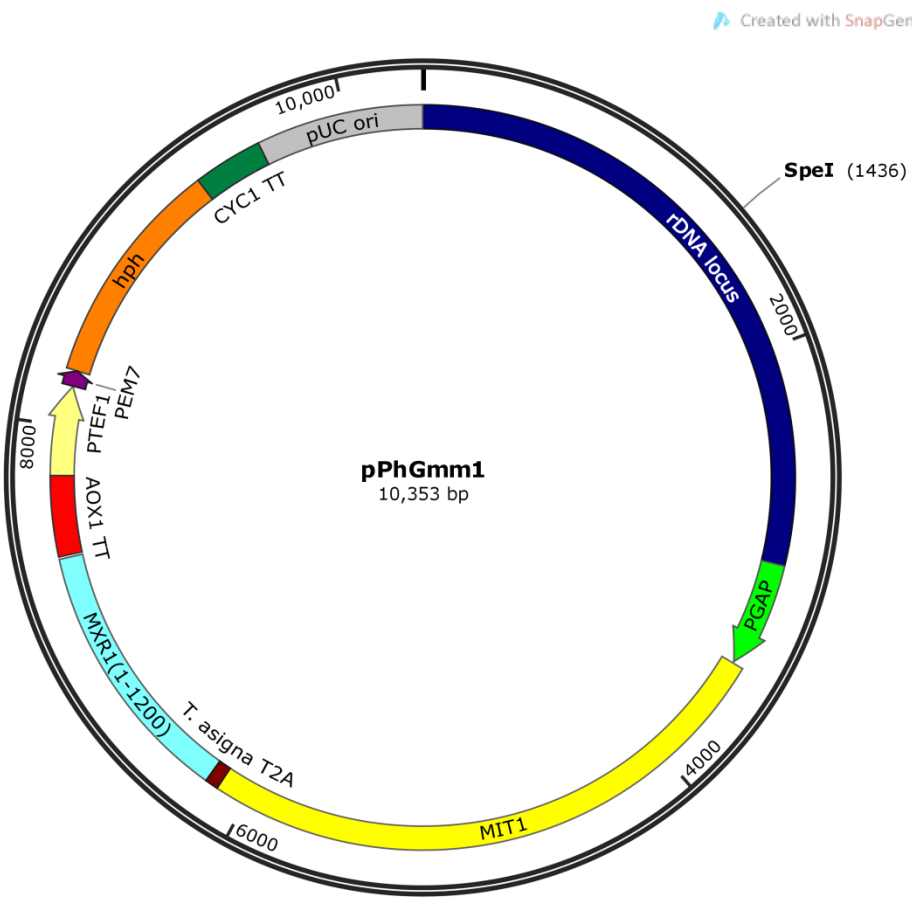


Figure 51: Vector map of pPhGmm1 for the simultaneous, strong constitutive expression of both *MIT1* and a truncated, functional form of *MXR1*.

6.4.1.3 Disruption of Mxr1p regulation through tagging with a C-terminal NLS

An alternative strategy to achieve true uncoupling of the function of Mxr1p from glucose repression, through attempting to disrupt the relocalisation of Mxr1p with a C-terminal nuclear localisation signal (NLS) was devised. NLS's, specifically those classified as "classic" nuclear import pathway, typically consist of one (monopartite) or two (bipartite) short sequences of positively charged amino acids on the protein surface (Dingwall and Laskey, 1991). Classic NLS's (cNLS) direct the transport of tagged proteins through the nuclear membrane by binding to karyopherin proteins that facilitate their import through nuclear pore complexes (Mafori et al, 2011). Analysis of the amino acid sequence of Mxr1p with the cNLS prediction software – cNLS Mapper (Kosugi et al, 2009) was unable to identify

any putative partite or bipartite NLS's, suggesting that its nuclear import in the absence of glucose is potentially mediated by an unrelated pathway. The addition of a cNLS could provide Mxr1p with an alternative, unregulated route into the nucleus, thus increasing nuclear levels of Mxr1p during glucose repression. The SV40 NLS from the Simian vacuolating virus (Kalderon et al, 1984) was selected to tag Mxr1p as it has previously been demonstrated to effectively direct the nuclear localisastion of proteins tagged at the C-terminus in *P. pastoris* (Weninger, Glieder and Vogl, 2015). The truncated form consisting of the N-terminal 400aa of Mxr1p (Mxr1t) was preferentially chosen to be expressed with a C-terminal SV40 NLS, with the aim of creating a minimal transcription factor for P_{AOX1}, with fewer potential regulatory regions that could be bound by antagonistic factors to inhibit its transport or function. To complement strains constitutively expressing *MIT1*, and to immediately test for glucose repression, pAVECRS was redesigned to also contain an expression cassette for Mxr1t-SV40 downstream of a region encoding its native promoter – identified by Lin-Cereghino et al (2006) (Figure 52).

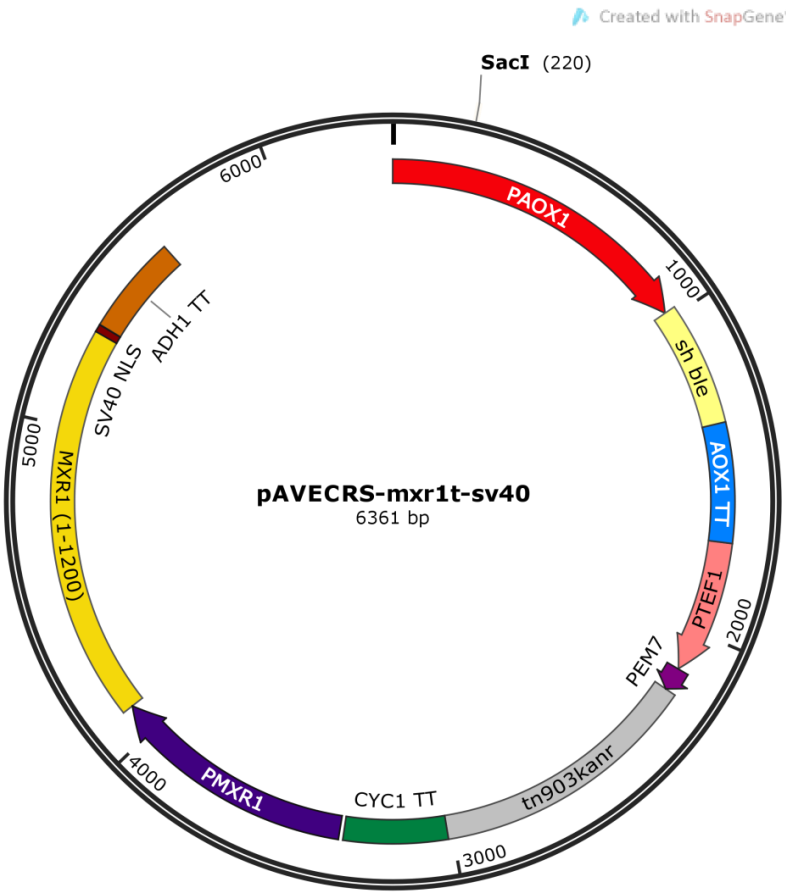


Figure 52: Vector map for pAVECRS-mxr1t-sv40, controlling the expression of both *sh ble* under P_{AOX1}, and a functional component of Mxr1p (Mxr1t) with a C-terminal SV40 NLS.

6.4.1.4 Cloning

pPCas9, assembled as described in section 6.3.2.1, was restricted with MssI to excise *CAS9-SV40* to produce the vector backbone for pPhGmit1 and pPhGmm1. In the case of pPhGmit1, *MIT1* was amplified from NRRL 11430 genomic DNA with the primers 99-GibGapmit1-F and 100-Gibmit1AOX-R, and inserted into Mss1-restricted pPCas9 by Gibson assembly. For pPhGmm1 *MIT1* was instead amplified with 99-GibGapmit1-F and 107-Gibmit1T2A to add a 3' Gibson end complementary to the downstream T2A, and *MXR1T* was amplified from NRRL 11430 cDNA with 109-GibT2Amxr1t-F and 110-Gibmxr1tAOX1-R (Figure 53).

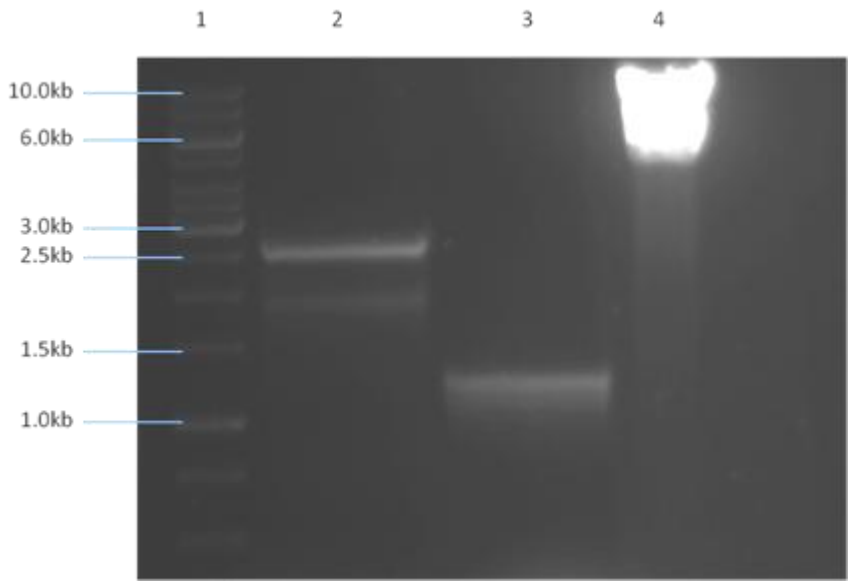


Figure 53: Generation of parts for the assembly of pPhGmm1. Lane 1 - 1.0kb DNA marker, 2 – PCR amplicon of the ~2.7kb *MIT1* gene with Gibson ends 3 - PCR amplicon the ~1.2kb MXR1T with Gibson ends, 4 – Restriction digest of pPCas9 with MSSI removing *CAS9-SV40*.

Gibson assembly of *MIT1*, *MXR1T* and 108-GibT2A, a synthesised 100bp oligonucleotide encoding the T2A sequence with Gibson ends for both *MIT1* and *MXR1T*, was conducted to insert the full *MIT1-T2A-MXR1T* sequence into MssI-restricted pPCas9. pPhGmit1 and pPhGmm1 were individually transformed into *E.coli* BioBlue and selected on 100µg/ml hygromycin b. The correct assembly of both vectors was confirmed by sequencing plasmids isolated from selected positive transformants with 111-5'GAP and 13-3'AOX1.

To insert an *MXR1T-SV40* expression cassette into pAVECRS, the vector was linearised with PciI to expose an independent site between the *CYC1* transcription terminator and the bacterial origin of replication. A fragment consisting of the region -755 - +1200 to the *MXR1* gene to contain *MXR1T*, its native promoter and compatible Gibson ends for insertion into pAVECRS was PCR amplified with 71-Gibmxr1p-F and 72-Gibmxr1sv40-R. 72-Gibmxr1sv40-R also adds the SV40 sequence and a stop codon immediately 3' of *MXR1T* in frame. As a transcription terminator for the new cassette the *ADH1* TT was PCR amplified from pG1-689-adht with the primers 73-Gibadh1tt-F and 74-Gibadh1tt-R. Both fragments were inserted simultaneously into the linearised pAVECRS backbone, transformed into *E. coli* BioBlue and selected on 30µg/ml kanamycin. The insertion of *MXR1T* was verified by a single restriction digest of plasmids isolated from selected positive transformants with LguI, which

cuts specifically within *MXR1T* (Figure 54). The correct configuration of the entire *MXR1T-SV40* expression cassette was also confirmed by sequencing plasmids tested as positive for containing *MXR1T* with the primers 75-mxr1sv40seq-F and 76-mxr1sv40-R.

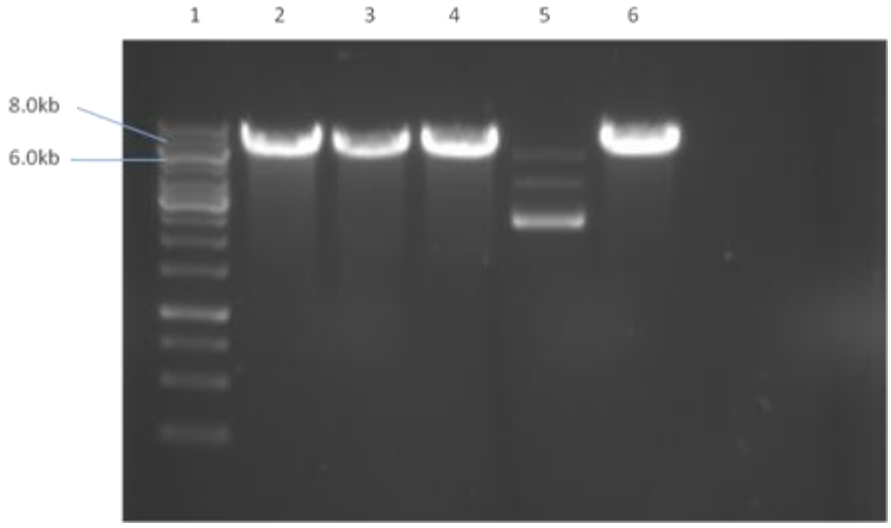


Figure 54: Diagnostic restriction digest of pAVECRS-mxr1t-sv40 with LguI. Lane 1 - 1.0kb DNA marker, 2-6 – Single restriction digest of plasmids from positive transformants with LguI. A single band ~6.0kb in size indicates the assembly of full pAVECRS-mxr1t-sv40.

6.4.2 Constitutive expression of *MIT1* abolishes its transcriptional repression in glucose-grown cells

pPhGmit1 and pPhGmm1 were linearised within the rDNA locus with SpeI and cloned into *P. pastoris* NRRL 11430 by electroporation to generate the strains: NGmit1 and NGmm1, respectively. To prevent the introduction of differences in gene dosage of transcription factors between strains as another variable, both NGmit1 and NGmm1 were verified to contain a single copy of the transformed vectors by qPCR of the genomic DNA of a selection of positive transformants with the primers 114-qhygb2-F and 115-qhygb2-R, which anneal to the *hph* selectable marker.

Firstly the GAP-driven expression of *MIT1* was tested to find if it provides similar quantities of *MIT1* mRNA transcript to the methanol-induced state even in glucose-grown cells. 5ml of liquid minimal

media containing 1% (w/v) glucose and 0.5% (v/v) methanol, having previously been demonstrated to cause repression of *AOX1*-based expression, was inoculated with colonies of either NRRL 11430 or NGmit1 to a starting OD_{600} of 0.1. Growth was maintained for ~16 hours to allow cells to reach exponential phase growth and *MIT1* expression was analysed by RT-qPCR of RNA isolated from each culture with the primers 116-qmit1-F and 117-qmit1-R. The expression of *MIT1* in both strains was compared against native *MIT1* levels during methanol induction by culturing NRRL 11430 under identical conditions in a minimal medium containing 0.5% (v/v) methanol, with subsequent RT-qPCR analysis (Figure 55).

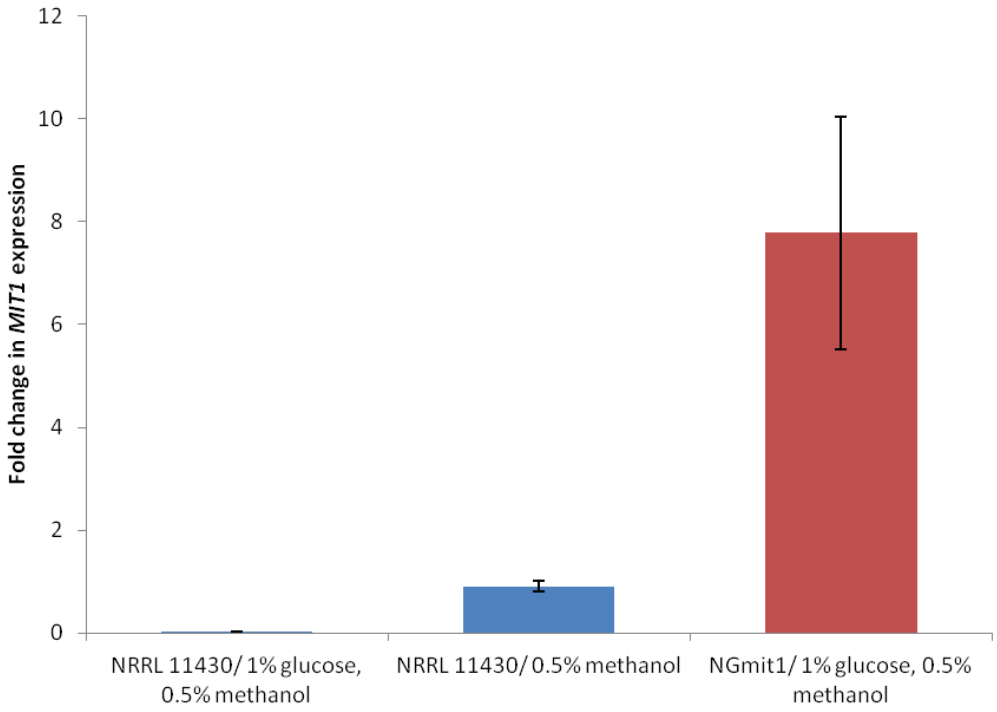


Figure 55: Fold change in expression of *MIT1* relative to *ACT1* of NGmit1 and NRRL 11430 following ~16 hours of growth in 1% (w/v) glucose and 0.5% (v/v) methanol, as well as methanol induced NRRL 11430 grown in 0.5% (v/v) methanol as a sole carbon source. Error bars represent the standard error of 3 biological replicates.

As expected the expression of *MIT1* in NRRL 11430 grown in media containing 1% (w/v) glucose was low, with a fold change of 0.02 relative to the expression of beta-actin, implying that its transcription was repressed in conditions eliciting glucose repression as has been previously reported (Prielhofer et al, 2015; Wang et al, 2016). Also consistent with previous reports on its expression profile was that

upregulation in *MIT1* transcription was also observed during methanol induction, with NRRL 11430 grown in 0.5% (v/v) methanol expressing *MIT1* at a ~45 fold higher level than cells grown in 1% (w/v) glucose and 0.5% (v/v) methanol. The substantial increase in *MIT1* transcription found in published results initially raised concerns that *GAP*-based expression would be insufficient to emulate methanol induced levels of *MIT1* in glucose-grown cells. However, following 16 hours of growth in 1% (w/v) glucose and 0.5% (v/v) methanol, *MIT1* was expressed over 7.5 fold higher on average in NGmit1 compared to native, methanol induced *MIT1* levels following a similar growth period. The results indicate that expression of a single copy of *MIT1* by P_{GAP} is not only sufficient to completely abolish its transcriptional repression during growth on mixed feeds of glucose and methanol, but also results in its overexpression. Nevertheless glucose repression experiments were continued with NGmit1 and NGmm1 with the caveat that tuning the strength of *MIT1* recombinant expression using a weaker promoter, to more closely resemble methanol induced levels, may have to be taken into future consideration if overexpression is found to be detrimental towards cell viability or productivity.

6.4.3 Glucose repression screening of NGmit1, NGmm1 and NGmit1 + Mxr1t-sv40

To test for the alleviation of glucose repression in strains constitutively expressing *MIT1* and *MXR1T*, NGmit1 and NGmm1 were cloned with pAVECRS, screened for copy number by qPCR with 93-qshble2-F and 94-qshble2-R and isolated as single copy integrants for the *sh ble* gene. To also test the alternate strategy of complementing the *MIT1* over-expressing strain with an NLS-tagged variant of Mxr1p, NGmit1 was also cloned with the modified pAVECRS-mxr1t-sv40 and screened to find a single copy integrant. The 3 experimental strains, each now containing a single copy of *sh ble* regulated by P_{AOX1}, were screened for their sensitivity to zeocin on minimal gradient agar plates containing either 2% (w/v) glucose (Figure 56), 0.5% (v/v) methanol/ 3% (w/v) sorbitol (Figure 57) or 1% (w/v) glucose/ 0.5% (v/v) methanol (Figure 58) as described in 2.4.3.

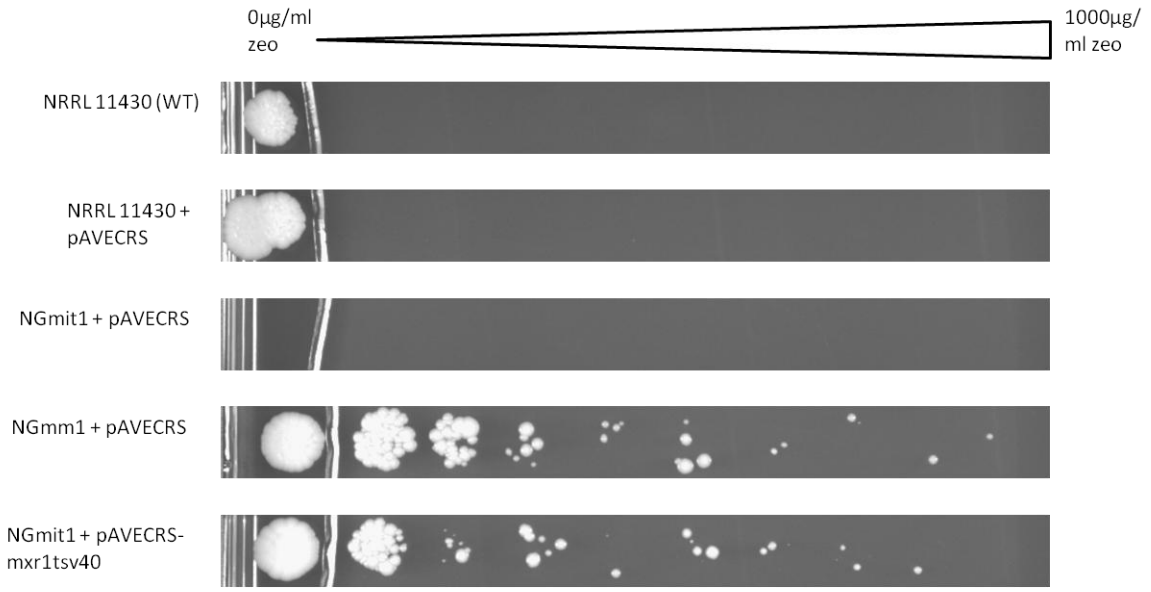


Figure 56: Sensitivity of the 3 experimental strains: NGmit1 (*MIT1*), NGmm1 (*MIT1, MXR1T*) and NGmit1 + mxr1t-sv40 (*MIT1, MXR1T-SV40*) – containing the pAVECRS screening vector to zeocin at concentrations up to 1000µg/ml on 2% (w/v) glucose. NRRL 11430 (WT) and NRRL 11430 +pAVECRS were simultaneously tested as negative control/ reference strains.

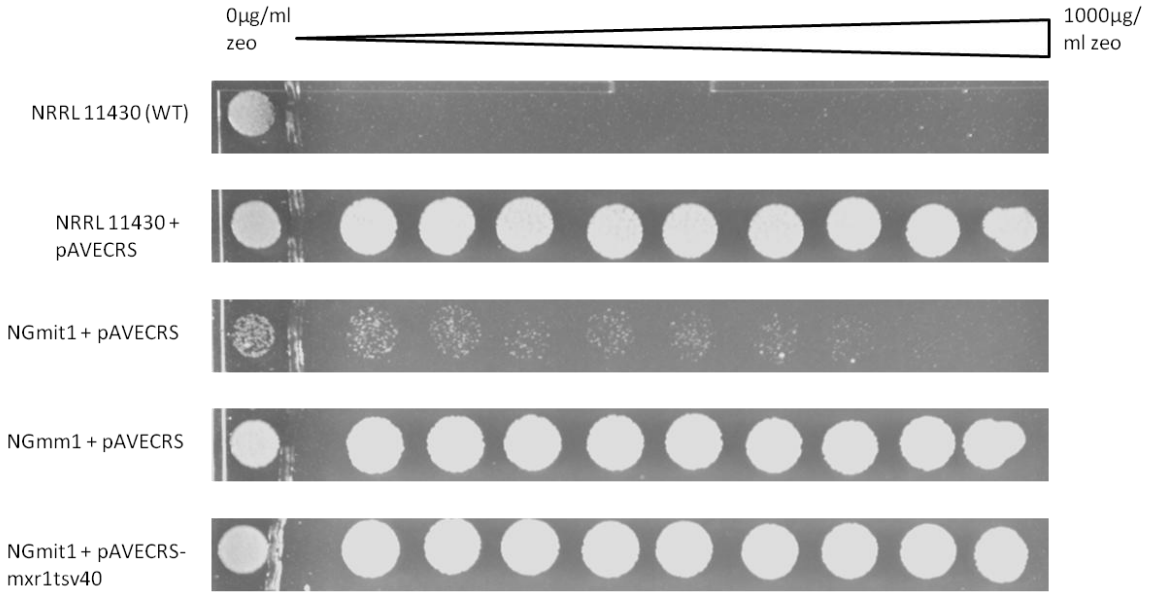


Figure 57: Sensitivity of the 3 experimental strains: NGmit1 (*MIT1*), NGmm1 (*MIT1, MXR1T*) and NGmit1 + mxr1t-sv40 (*MIT1, MXR1T-SV40*) – containing the pAVECRS screening vector to zeocin at concentrations up to 1000µg/ml on 0.5% (V/v) methanol and 3% (w/v) sorbitol. NRRL 11430 (WT) and NRRL 11430 +pAVECRS were simultaneously tested as negative control/ reference strains.

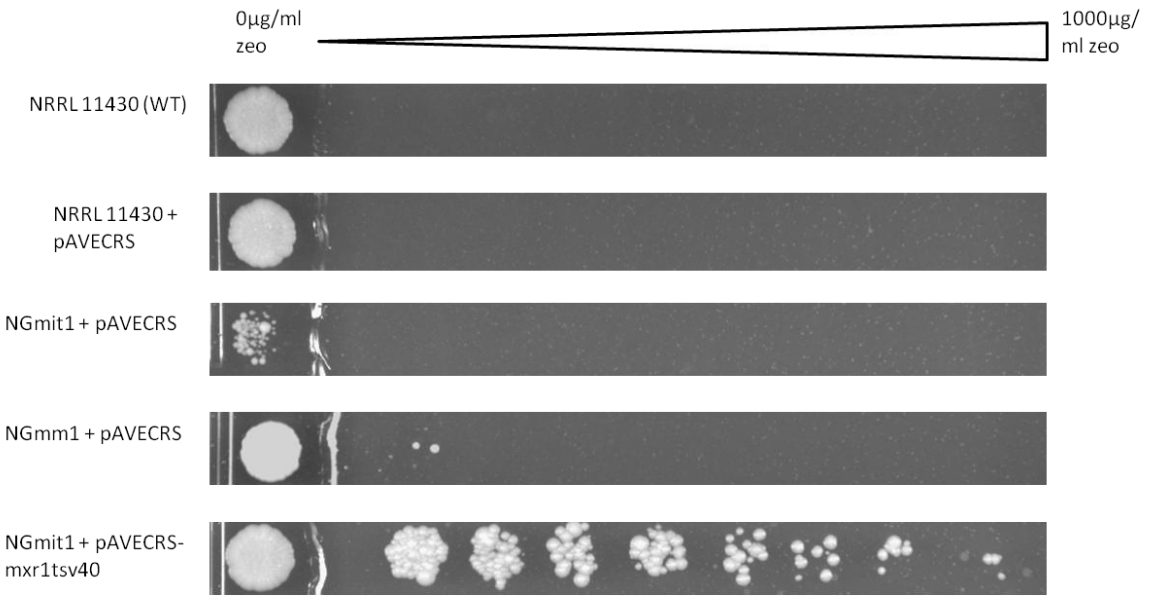


Figure 58: Sensitivity of the 3 experimental strains: NGmit1 (*MIT1*), NGmm1 (*MIT1, MXR1T*) and NGmit1 + mxr1t-sv40 (*MIT1, MXR1T-SV40*) – containing the pAVECRS screening vector to zeocin at concentrations up to 1000µg/ml on 0.5% (V/v) methanol and 1% (w/v) glucose. NRRL 11430 (WT) and NRRL 11430 +pAVECRS were simultaneously tested as negative control/ reference strains.

Surprisingly NGmit1 showed little to no growth in each of the tested media types, even within the control region on plates containing no zeocin. Some growth was observed across the zeocin gradient in inducing conditions, suggesting that *AOX1* expression was still occurring. However NGmit1 was capable of forming regular colonies on standard YPD agar and grow in liquid media so the overexpression of *MIT1* itself was not assumed to be lethal. It was noted that NGmit1 grew to a lower OD_{600} in all of the starter cultures for the glucose repression screens, suggesting that, though viable, NGmit1 appeared to suffer impaired growth in minimal media. Since the glucose repression screen associates P_{AOX1}-driven expression not solely by overall protein production but also cell growth in response to excessive quantities of antibiotic it could be possible that, despite expressing *sh ble*, the low growth rate of NGmit1 coupled with the cellular stress incurred by high zeocin levels still resulted in a loss of viability. The antibiotic within gradient plates also begins to diffuse laterally over extended periods of time, which could present a further disadvantage to slow growing strains. If the time required for NGmit1 to establish colonies on minimal agar exceeds the period in which the zeocin gradient begins to normalise it could provide an explanation as to why the strain was unable to grow even at lower concentrations. It was therefore concluded that the glucose repression screen in

its current form is not able to accurately determine the transcriptional activation of P_{AOX1} in NGmit1, or for any slow growing recombinant strains for that matter.

Despite also overexpressing *MIT1*, NGmm1 exhibited normal growth and colony formation in regions containing no zeocin. Very minimal derepression of P_{AOX1} could also be observed in NGmm1 growing on 0.5% (v/v) methanol/ 1% (w/v) glucose; forming 2 colonies on the lowest zeocin concentration region tested. *sh ble* expression in NGmm1 was significantly higher when grown in 2% (w/v) glucose, indicating that "leaky" expression, in which P_{AOX1} is partially induced even in the absence of methanol, occurs when *MIT1* or *MXR1T* are overexpressed. The strength of P_{GAP} is increased when cells are grown on glucose (Waterham et al, 1997), and could be the reason why *sh ble* expression was much higher for NGmm1 when grown in the standard concentration for yeast culture of 2% (w/v) glucose as opposed to the glucose/ methanol mixed feed, which only contained 1% (w/v) glucose. However this would be based on the assumption that intracellular levels of glucose vary as the concentration of extracellular glucose is increased past 1% (w/v).

Of each of the three novel strains tested NGmit1 containing *MXR1T-SV40* performed the best, exhibiting the highest resistance to zeocin when grown in 0.5% (v/v) methanol/ 1% (w/v) glucose and similar, if not less, leaky expression to NGmm1 when grown on glucose without methanol. The increase in colony frequency when grown in methanol/ glucose over solely glucose indicates that both derepression and partial methanol induction of P_{AOX1} can be effected by NGmit1 + *MXR1T-SV40* in the presence of 1% (w/v) glucose. The results suggest that, rather than being able to freely bind and activate *AOX1* transcription simply when present in the nucleus, either one or both transcription factors have individual activation states determined by the presence of methanol in the cell.

To further quantify, and establish in the case of NGmit1, the degree of glucose repression alleviation in each strain, samples of each culture in 0.5% (v/v) methanol/ 1% (w/v) glucose were harvested immediately prior to plating onto zeocin gradients for RNA purification. RT-qPCR for s*h ble* of RNA from each sample, in addition to RNA from NRRL 11430 + pAVECRS grown in 0.5% (v/v) methanol/ 3% (w/v) sorbitol as a reference for full P_{AOX1} activation, was conducted with the primers: 93-qshble2-F and 94-qshble2-R (Figure 59).

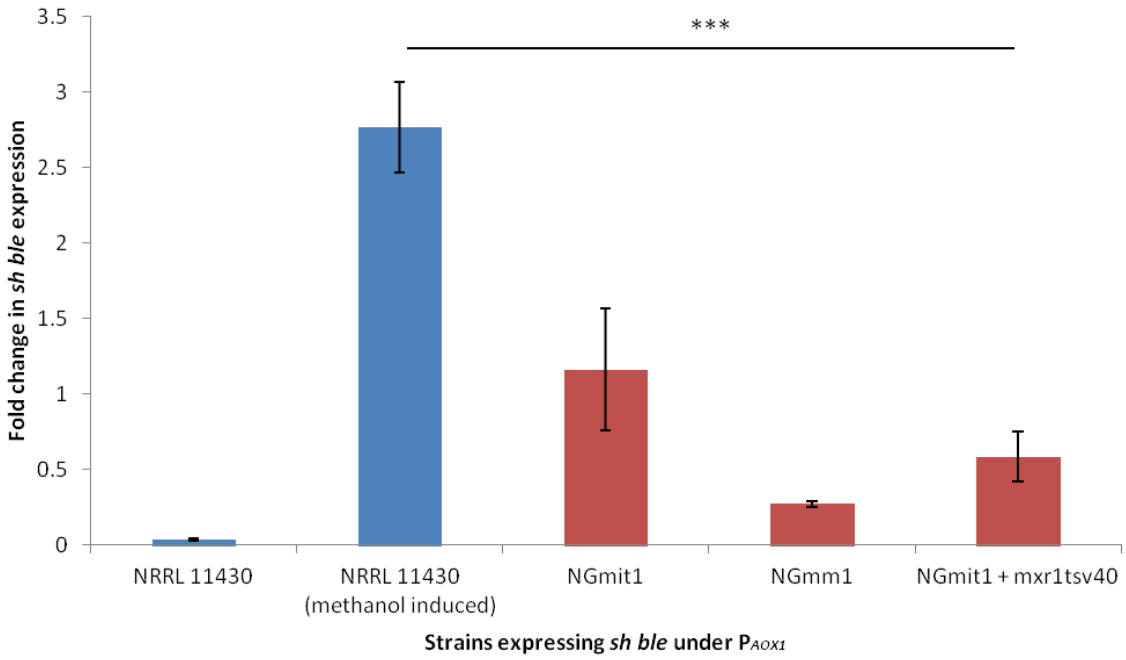


Figure 59: Fold change in expression of *sh ble* expression between NRRL 11430 (WT), NGmit1 (*MIT1*), NGmm1 (*MIT1, MXR1T*) and NGmit1 + mxr1t-sv40 (*MIT1, MXR1T-SV40*) – containing pAVECRS following ~16 hours of growth in 0.5% (v/v) methanol, 1% (w/v) glucose, prior to screening for catabolite repression. *sh ble* expression from NRRL 11430 + pAVECRS grown in 0.5% (v/v) methanol, 3% (w/v) sorbitol was also tested as a positive control for full P_{AOX1} activation within the set conditions. Expression values are calculated as a fold change in relation to the expression of *ACT1*. Error bars represent the standard error of 3 biological replicates. The results of an ANOVA test for the homogeneity of mean expression between the sample sets specified are displayed - ns: p>0.05, *: p<0.05, **: p<0.01, ***: p<0.001

The data from the RT-qPCR confirms a number of observations made during the initial glucose repression screen, namely that each of the novel strains were repressed to a lesser extent than the native strain but that full derepression of P_{AOX1} did not occur when cells were grown in minimal methanol medium supplemented with 1% (w/v) glucose. Despite growing poorly on zeocin gradient plates during the initial screen NGmit1 achieved the highest mean transcription of sh ble under glucose repressing conditions, followed by NGmit1 expressing *MXR1T-SV40*, although a Games-Howell post hoc test comparing the two found no statistically significant difference between the overall expression levels. NGmm1 was also confirmed to exhibit the weakest derepression of P_{AOX1} in the glucose/ methanol mixed medium. Considering that it should be expressing equal quantities of *MIT1* to both NGmit1 and NGmit1 + mxr1t-sv40, and that increased derepression occurs at higher glucose concentrations, it was concluded that NGmm1 is inefficiently synthesising Mit1 and Mxr1t,

thus requiring stronger promoter activities to match the derepression observed in the other strains. This could have been brought about by inefficient ribosome skipping at the T2A sequence during translation, resulting in a portion of translated Mit1 and Mxr1t forming a non-functional hybrid rather than two discrete proteins. Inefficient ribosome skipping has previously been reported for the use of T2A sequences in *P. pastoris*, though the severity of the effect and the ensuing ratio of protein fusions to mature protein depend on the sequences of the respective genes as well as their order within the polycistronic cassette (Weninger, Glieder and Vogl, 2015).

6.4.4 Production of synthetic human lysozyme T70A by NGmit1 and NGmm1 in mixed glucose and methanol media

To determine how the results observed in the glucose repression screen translate to the secreted production of recombinant proteins in glucose-supplemented expression media, microexpression trials were designed for NGmit1 and NGmm1. HuL T70A was selected as a marker to test for productivity due to the rapid screening methods available to assay its activity and of the 2 main proteins tested within this project, the other being glucose oxidase, the secretion of HuL T70A would not compromise a glucose-based medium. Previous expression studies (5.4.3) also showed that the expression of a single copy of *HUL T70A* does not induce the excessive levels of ER stress found in previous studies. As both strains are immediately compatible with pPICZ vectors pPICZm-T70A was linearised within the *AOX1* locus to clone into NGmit1 and NGmm1 by electroporation. Positive transformants were selected with 100µg/ml zeocin and screened for single copy integrants by qPCR of genomic DNA with 83-qsynHuL3-F/ 84-qsynHuL3-R as previously described. The resulting strains were designated as NGmit1-AαT70A and NGmm1-AαT70A

As the previous screen showed - concentrations from 1% (w/v) glucose in the growth medium continue to significantly repress *AOX1*-based expression in all strains so newly defined methanol expression media, supplemented with glucose to a final concentration of either 0.25% (w/v) or 0.5% (w/v), was composed to test the performance of NGmit1-AαT70A and NGmm1-AαT70A. Expression trials for both strains, in addition to native NAαT70A were conducted in 24 well microtiter plates from a starting OD_{600} of 1.0 in BMMY, BMMY + 0.25% (w/v) glucose or BMMY + 0.5% (w/v) glucose over a 48 hour period. To maintain glucose levels within each of the mixed media types, 20% (w/v) glucose

was added in combination with methanol up to the specified final concentration of the original medium at 24 hour intervals. HuL activity was assayed in culture supernatants to quantify the volumetric and specific productivities of HuL T70A for each strain within the range of glucose concentrations tested (Figure 60).

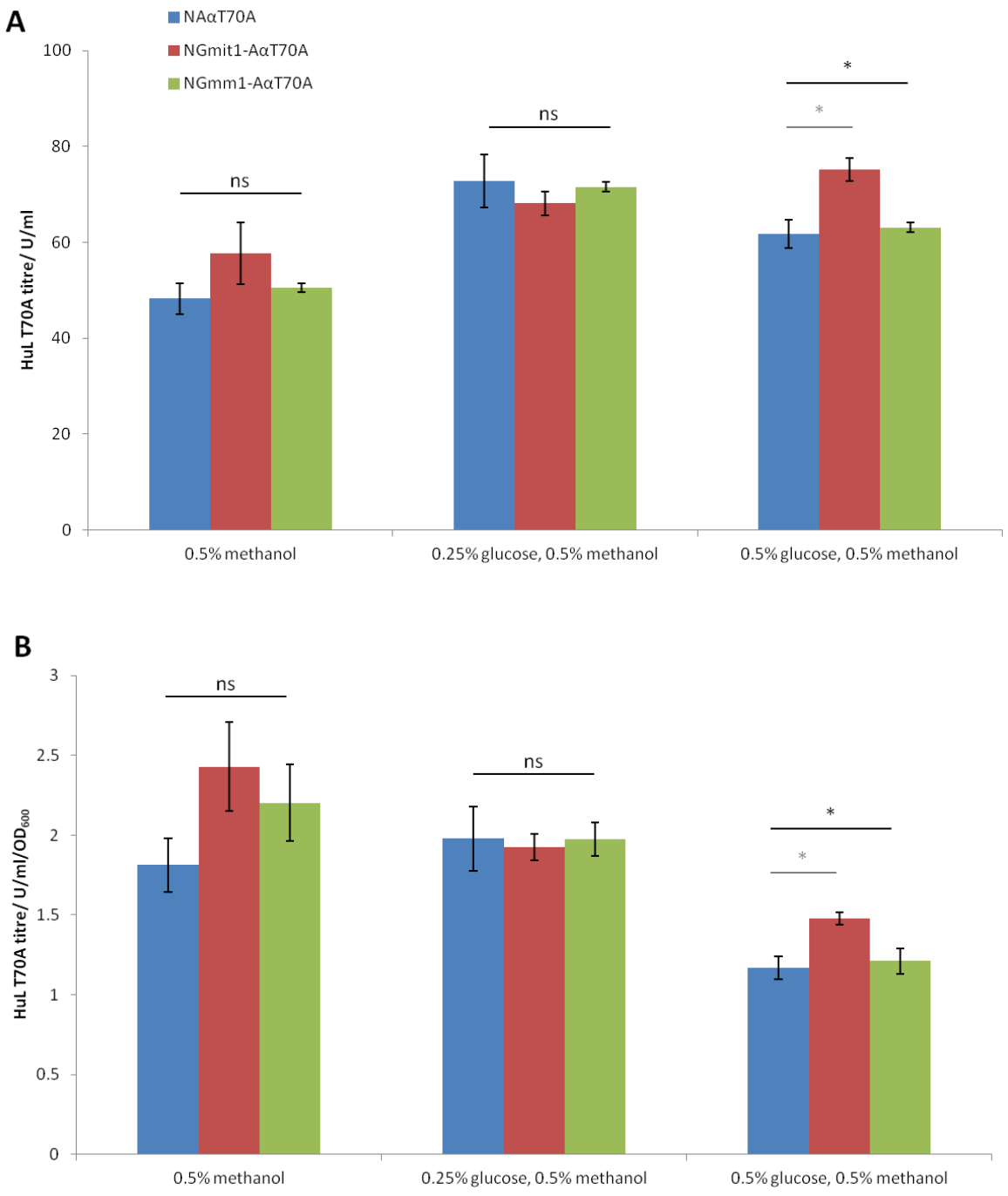


Figure 60: Comparison of the volumetric productivity (A) and the specific productivity (B) of secreted HuL T70A by NAαT70A (native), NGmit1-AαT70A (*MIT1*) and NGmm1-AαT70A (*MIT1*, *MXR1T*) in BMMY (0.5% (v/v) methanol supplemented with increasing concentrations of glucose (w/v). Error bars represent the standard error of 4 biological replicates. The results of ANOVA tests (black line) between the mean values of each strain with relevant Games-Howell post hoc analyses (grey line) are displayed - ns: $p>0.05$, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$

The impaired growth of NGmit1 observed in minimal media was not evident in any of the BMMY-based media tested within the experiment, as it achieved similar final cell densities to the native strain under each condition; suggesting that no loss in viability occurs during growth in rich media. In standard BMMY containing no glucose NGmit1-AαT70A slightly outperformed NAαT70A and produced ~20% more active HuL T70A on average. However an ANOVA between NAαT70A, NGmit1-AαT70A and NGmm1-AαT70A could not determine a statistically significant difference between the mean volumetric or specific productivities of each, most likely due to the wider distribution of HuL activities recorded between the biological replicates specifically for expression on 0.5% (v/v) methanol. The addition of glucose at a final concentration of 0.25% (w/v) caused a concurrent increase in the overall productivity for all three strains, signifying that the utilisation of glucose was sufficient to stimulate higher protein production through increasing cellular growth. No difference in productivity could be observed between any of the tested strains, suggesting that 0.25% (w/v) glucose is not fully repressive or doesn't prolong repression of *AOX1* expression, as has been previously demonstrated (Paulová et al, 2012). An alternative possibility, and an initial concern, was that the growth of cells in glucose-fed cultures cannot be regulated effectively in micro expression trials. More rapidly growing cultures in methanol/glucose mixed media would be able to deplete the available glucose to limiting concentrations earlier, reducing the period in which CCR is maintained during the expression. However glucose repression became evident when the final glucose concentration was increased to 0.5% (w/v), causing a decrease in the volumetric productivity for NAαT70A and NGmm1-AαT70A from expression in 0.5% (v/v) methanol/ 0.25% (w/v) glucose. It was reasoned that the addition of 0.5% (w/v) glucose at 24 hour intervals was able to create a large enough window of *AOX1* repression before being consumed to produce a noticeable effect on total production of HuL T70A after 48 hours. However NGmit1-AαT70A remained unaffected and even showed a slight increase in its volumetric productivity. Again, NGmit1-AαT70A produced ~20% more HuL-T70A than the native strain per ml of culture supernatant, which was found through a post hoc analysis from a one way ANOVA to be statistically significant within the 95% confidence level. The difference was found to be more pronounced when examining specific productivity as NGmit1-AαT70A exhibited over a 25% increase in the titre of HuL T70A, suggesting that an increase in cellular production, rather than improved growth in 0.5% (v/v) methanol/0.5% (w/v) glucose, was the underlying cause. The results indicate that the *GAP*-based expression of *MIT1* alleviates glucose repression of P_{AOX1} at lower concentrations, thereby increasing the threshold for the total glucose useable in culture beyond which protein production is negatively affected.

In contrast NGmm1-AαT70A displayed negligible differences in productivity from NAαT70A in all the media types tested, suggesting that it was experiencing native *AOX1* regulation in response to methanol and glucose. This reaffirmed previous assumptions that the translation and synthesis of Mit1 was more inefficient when expressed from the polycistronic cassette with *MXR1T*.

6.5 Discussion

The study began with the design and successful implementation of a screen to measure *AOX1* promoter activity in cells grown on glucose or methanol, based on the P_{AOX1}-driven expression of the zeocin resistance marker. A screen for the activation of methanol utilisation in the presence of glucose has been previously developed for methylotrophic yeasts and involves growing cells on a medium containing methanol and 2-deoxy-D-glucose; a glucose analogue capable of inducing CCR of Mut pathway genes (Stasyk et al, 2004). 2-deoxy-D-glucose cannot be metabolised and so cell survival can only be ensured through co-utilisation of the available methanol. However the screen developed within this study directly assays the situational promoter strength of P_{AOX1}, rather than testing for the full activation of the Mut pathway, and is therefore more suited to identifying strains carrying out high levels of *AOX1*-based transgene expression. The low setup time and the ability to test multiple strains simultaneously give the screen a high degree of versatility and in future could be applied to larger experiments examining libraries of variants, with examples including directed evolution or combinatorial approaches looking to improve the function and optimise the expression of the transcriptional activators tested here. Furthermore, full repression of P_{AOX1} in strains containing pAVECRS was found to result in cell death even at lower concentrations of zeocin. The screen could therefore be adapted into a selection method for the derepression of *AOX1* with a low occurrence of false positives, and combined with random mutagenesis experiments to identify novel mutants with impaired glucose repression of P_{AOX1}.

In order to undertake a multiplex gene disruption strategy to develop strains with increased alleviation of glucose repression over the single knockout mutants previously described in literature CRISPR-Cas9 gene editing in *P. pastoris* was attempted but was ultimately unsuccessful. However, since the conclusion of this study, a CRISPR-Cas9 system with high targeting efficiency has been validated in *P. pastoris* (Weninger et al, 2015B). The authors tested a combinatorial library consisting

of 95 constructs varying in constitutive and inducible promoters, gRNA expression cassette types, gRNA scaffold sequences and codon usage biases for the *CAS9* sequence (Weninger et al, 2015B). Of the constructs screened only ~6% yielded efficiently targeted CRISPR-Cas9 activity, concluding that *P. pastoris* is not as amenable to CRISPR-Cas9 gene editing as other established species, and that more precise optimisation is required to engineer an effective system within the organism. Interestingly the constructs that functioned correctly bear similarities to the strategies used within this study, including the expression of human optimised, SV40 tagged *CAS9* by P_{GAP}, and the use of the HH and HDV ribozymes in conjunction with RNA pol II *P. pastoris* constitutive promoters for the expression of gRNAs. Where the constructs differ are in the use of the *HTB* or *HTX* promoter to drive gRNA expression, the use of an alternative scaffold sequence for the gRNA (Jinek et al, 2012) and the transient expression of both *CAS9* and gRNAs on an episomal, autonomously replicating plasmid. The failure of the CRISPR-Cas9 system in this study can therefore be pinpointed to the poor expression of functional gRNAs, either through suboptimal transcription rates through the use of the G1 promoter or the transient expression method, or the scaffold sequence for the gRNA failing to form the correct secondary structure to dock with Cas9. The development of proficient CRISPR-Cas9 editing is an invaluable addition to the *P. pastoris* molecular tool kit and its associated publication can inform the modification of the parts made within this study to replicate the working system for future experiments targeting glucose repression.

In the absence of a system for enabling simple and rapid multiple gene deletions the study moved towards uncoupling transcription enhancers of *AOX1* from their inhibition in glucose-grown cells, with focus on the methanol induced transcription factor – Mit1. Expressing *MIT1* constitutively under the *GAP* promoter was capable of not only abolishing its transcriptional repression during growth on a mixed feed of glucose and methanol but also increasing expression beyond its native level in methanol-grown cells. This sole modification significantly alleviated glucose repression of *AOX1*-based expression and enables the use of higher concentrations of glucose with methanol in expression media that otherwise begin to be detrimental to methanol-induced protein production in the native strain. The expression of both *MIT1* and *MXR1T* in a polycistronic cassette was intended to improve the efficiency of transformation and expression of both genes but instead resulted in less derepression in comparison to the sole expression of *MIT1*, despite the overexpression of *MXR1* having been previously shown to reduce glucose repression (Takagi et al, 2012). This has made the assessment of the effect of co-overexpressing both Mit1 and Mxr1p somewhat difficult. The expected alleviation of glucose repression was restored when expressing *MIT1* and *MXR1T*

separately, as observed in the strain NGmit1 + mxr1t-sv40. Employing polycistronic cassettes using 2A peptides would still be very beneficial for this type of expression in *P. pastoris* through scaling down the number of expression cassettes or vectors required and so optimisation of the translation efficiency of Mit1 and Mxr1t with T2A should be considered. Potential strategies include switching the order of each gene within the expression cassette or the addition of codons encoding a glycine-serine-glycine spacer at the N-terminus of the T2A peptide, which has been shown to improve cleavage efficiency (Szymczak et al, 2004) (Holst et al, 2006).

Another promising result that warrants further investigation was the co-expression of a gene encoding a truncated form of Mxr1p with a C-terminal SV40 NLS, with *MIT1*. Whilst exhibiting similar levels of P_{AOX1}-driven expression to strains overexpressing only *MIT1* when grown in 0.5% (v/v) methanol/ 1% (w/v) glucose, NGmit1 + mxr1t-sv40 exhibited standard growth on minimal media during glucose repression screening whereas *MIT1* overexpressing strains grew poorly. The results highlight a phenotypic difference between the two strains that implies that the expression of *MXR1T-SV40* is providing some effect. However the basis for the difference in growth on minimal media between each strain can only be speculated at. As each of them have been shown to regulate methanol inducible promoters differently, it could be possible that increasing the expression of only one of the master regulators of methanol metabolism results in an unbalanced activation of a subset of the Mut pathway that may potentially interfere with other metabolic or biosynthetic pathways within the cell. Whilst this has not been confirmed previously, the overexpression of *MXR1* under P_{AOX1} was found to be lethal in methanol-grown cells (Lin-Cereghino et al, 2006).

Also worthy of note is that, while partial P_{AOX1} activation in the presence of repressive quantities of glucose was exhibited, complete abolishment of glucose repression was not observed for any of the newly engineered strains. This could implicate the involvement of other interacting factors that govern the activity of the master regulators of the Mut pathway and suggests that the full picture for CCR of P_{AOX1} is yet to be fully understood, but may be revealed by looking at highly resistant variants. In order to shift methanol inducible expression systems in *P. pastoris* away from the reliance on methanol as a growth substrate, while still maintaining the characteristically high strength of P_{AOX1}, resistance to repression at final glucose concentrations up to 2% (w/v), the standard used for yeast growth, would be needed. Nevertheless the partial derepression presented in this study provides a starting point for optimisation strategies and strain improvement. Since the major form of regulation of Mit1 appears to be transcriptional, tuning its expression by testing a range of different strength

constitutive promoters could possibly alleviate glucose repression further. Evaluating the effectiveness of the SV40 NLS in targeting Mxr1t to the nucleus through the generation of fluorescent reporter fusion constructs would also be pertinent, especially since the mechanism by which Mxr1p is localised in the cytoplasm, and whether nuclear retention is the only condition required to eliminate the inhibition of Mxr1p during glucose-growth remains unknown. The constitutive expression of *MIT1* and *MXR1T* confers some resistance in cells cloned with pAVECRS to zeocin under glucose repression; a phenotype that can be targeted and sequentially improved through multiple rounds of mutation with antibiotic selection. Strains selected in this fashion would hypothetically accumulate mutations disrupting genes involved in glucose repression of P_{AOX1}, increasing the compatibility of *AOX1*-based transgene expression with glucose growth while also providing more insight into the factors mediating glucose repression. Finally the results of the micro expression trials presented in this study have aided in defining the range of glucose content that is optimal for protein production in *MIT1* which can begin to inform successive scale-up experiments. Testing the performance of NGmit1 and future strains in which Mxr1p function is uncoupled from glucose regulation in chemostat cultures is not only key in assessing their industrial relevance but also enables the precise control of growth substrate concentrations and biomass accumulation. Cultures maintained at steady state growth will be unable to deplete available glucose or methanol, causing the inherent fluctuations in carbon source concentrations between feeding intervals during rapid growth in batch trials. The initial results for the expression of HuL T70A in NGmit1 shows promise that sustaining constant rates of *AOX1*-based expression by preventing methanol and glucose concentrations dropping to limiting levels throughout expression, NGmit1 will further outperform both native *P. pastoris* at higher glucose: methanol ratios and for standard fermentation processes using methanol as the sole carbon source.

7. Final discussion and conclusions

7.1 The investigation of clonal variation and current research landscape

This project aimed to take a fundamental approach in identifying and evaluating the contribution of key factors that have been flagged as potential sources of variation in productivity within the *Pichia* system. The artificial environments and processes that *P. pastoris* is subjected to during transformation and transgene expression were first examined, identifying the widespread use of zeocin as a selective antibiotic as an initial focus of study. The increased mutagenic background caused by zeocin, even in resistant human cell lines, led to the hypothesis that exposure to zeocin during the formative stages of colony growth post-transformation would result in significant phenotypic divergence between positive transformants. However a comprehensive study examining the secreted production of four different recombinant proteins across two different host strains found that the standard application of zeocin in recombinant clone selection and colony growth has no effect on clonal variation. On the contrary the maintenance of a selective pressure, using either zeocin resistance or HIS4 complementation during colony growth, resulted in the formation of clonal groups with reduced variability in productivity. The mechanism behind this phenomenon, and the variation in its effect depending on the recombinant protein being produced could not be elucidated. Analysis of transgene copy numbers in clonal samples selected with zeocin however revealed a tighter distribution of copy numbers occurred within clones grown in the presence of zeocin, suggesting that zeocin may selectively enrich for specific copy numbers at different concentrations. However this pattern was not repeated in clonal populations utilising the alternative selection method while still reducing clonal variation and so either contradicts this assumption or supports that the phenomenon is unique to zeocin selection. As a result the conclusions made from the initial study were that, whilst shown to be mutagenic, the use of zeocin at the concentrations and exposure times used in the selection of *P. pastoris* transformants does not contribute to clonal variation, and that selection in itself actually confers an increased consistency in protein production within clonal populations.

Variation between recombinant clones in factors that have been previously linked to productivity, including transgene transcription, *HAC1* expression and growth were also explored. Whilst considerable variation was found within the sample of clones tested for each of the aforementioned

factors, only changes in the rate and total growth of cells in culture correlated weakly with recombinant protein secretion. The positive relationship between growth and productivity in *P. pastoris* has been covered extensively, having been found in a number of independent studies (Schenk et al, 2008; Love et al, 2010; Buchetics et al, 2011; Rebnegger et al. 2014). More specifically productivity has been demonstrated to be cell cycle dependent as, during faster growth and high protein secretion, a shift in the distribution of cell cycle phases of cells in culture towards G2 and the mitotic stage occurs (Buchetics et al, 2011). This research has recently been taken further to discover that a large proportion of recombinant protein secretion takes place at the bud tip of replicating cells during the G2 phase, confirming that productivity in *P. pastoris* is strictly coupled to active growth (Puxbaum, Gasser and Mattanovich, 2016). Indeed the higher producing subset of clones that were tested also displayed a small but significant shift in the distribution of their growth rates, with an increase in the median growth rate by a modest 4% over the low secreting clonal group. Whilst the previously mentioned studies varied culture growth through the control of specific growth rate in chemostat culture, the findings from this study also suggest that differences in the individual growth rates of clones in batch cultures provide a slight predisposition towards recombinant protein secretion. However the weakness in the correlation suggests that growth variation forms only a minor facet of the collection of positive traits improving the ability of individual clones to secrete recombinant proteins, the rest of which may prove challenging not only to identify but also to enrich in clonal populations. The reduction in the strength of the correlation found in previous studies (Rebnegger et al, 2014) suggests that the aspect of cell growth being studied within these experiments differs from other publications. In contrast to chemostat cultures that tightly control growth rate by increasing the rate of nutrient supplementation, the variation in growth observed in the batch trials of this study reflect the individual propensities of clones for transporting and metabolising the limiting substrate in culture. The results suggest the likelihood that variability in metabolism rates for growth substrates arises within generations grown from homogenous cell lines, and that this may contribute to individual performance of clones but that, ultimately, this variation can be controlled well in scale-up processes that maintain optimal specific growth rates for protein production. Nevertheless the link between growth rate and productivity has introduced another target for bioprocess development in the *Pichia* system in previous studies (Buchetics et al, 2011) and will most likely be investigated in future strain improvement experiments (Puxbaum, Gasser and Mattanovich, 2016).

At present the genetic instability in the integration and maintenance of multi-copy constructs within genomic loci remains the most influential, known factor affecting variation between *P. pastoris* clones. In fact research aimed at reducing the occurrence of additional recombination events between heterologous DNA sequences after transformation has made great strides in improving the consistency of transformants as well as their performance in longer fermentations. Notable examples include the discovery of the 5S rRNA gene locus, also used within this study (rDNA), as a better integration site for expression vectors to support their retention in the host genome (Steinborn et al, 2006; Marx et al, 2009). The rDNA locus is repeated 16 times across all chromosomes in the *P. pastoris* genome (De Schutter et al, 2009), presenting independent, spatially distant sites for vector integration such that the resulting multicopy strains are not necessarily generated by the head to tail multimerisation of vector sequence within a single locus; an orientation that is more prone to genetic instability. More recently a large-scale analysis of *AOX1* vector integration in *P. pastoris* screened 845 transformants producing GFP for recombinant productivity, copy number and integrated DNA sequence revealed several subgroups of copy number variants, transformants integrating vector copies in a head to head or tail to tail orientation as opposed to the expected head to tail configuration, Mut phenotype variation and false positives that had undergone a second recombination event to lose the expression vector – all of which produce a marked effect on productivity (Schwarzhans et al, 2016). By simply replacing the *AOX1* transcription terminator within the GFP expression cassette with the *S. cerevisiae CYC1* terminator, leaving the *AOX1* promoter as the sole region with any homology to the *P. pastoris* genome, the instances of false positives were significantly reduced to 0 out of 120 tested transformants (Schwarzhans et al, 2016). The results highlight the unsettling reality that mainstream transformation methods for *P. pastoris* produce a massive heterogeneity in the recombinant genotype between transformed cells, and the continued importance of rigorous screening of clones as part of the upstream process. Obvious targets that arise for the reduction of clonal variation by improving genetic stability would be the non-homologous end joining (NHEJ) and the homologous recombination (HR) pathways, both of which mediate the processes causing non-specific vector integration and copy number loss from loop-out recombination (Zhu et al, 2009B; Aw and Polizzi, 2013). As such the inactivation of NHEJ in *P. pastoris* has already been addressed in a study which deleted the *KU70* homologue, responsible in part for the initiation of the pathway, and found that the targeting of integrating vectors into the intended genomic locations occurred at a much higher efficiency whilst transgene copy numbers up to 7 could be maintained in strains over a 480 hour culture period. The lack of published results and also

previous work on the subject (Aw, 2012) suggests that a similar approach is not possible for limiting homologous recombination frequencies in *P. pastoris*, and that the complete disruption of the HR machinery is lethal. This is not entirely unexpected as homologous recombination is the dominant repair pathway for DNA damage in yeast (Sonoda et al, 2006) and its complete loss in and of itself is also undesirable as it forms the key mechanism for the integration and amplification of expression vectors in the genome (Higgins and Cregg, 1998; Sunga, Tolstorukov, and Cregg, 2008). However the recent development of next generation synthetic biology tools for *P. pastoris*, including inducible promoter libraries (Hartner et al, 2008) and a CRISPR-Cas9 system (Weninger et al, 2016), could be set to modernise pathway regulation in this system which, up until recently, has been limited to the binary decision of either knocking out or overexpressing target genes. We envisage that synthetic regulatory circuits, enabled by technologies such as CRISPR or TALENs already being used in yeast (Blount, Weenink and Ellis, 2012; Jusiak et al, 2016), will eventually be applied to fine tune gene expression in *P. pastoris* and overlay on/off switches to cellular processes affecting clonal variation or even productivity in future strain improvement. Since homologous recombination and end joining can be both beneficial or disruptive depending on the stage of the *Pichia* system process, these would make prime targets, and could be used to engineer platform strains that maintain or upregulate HR during transformation and vector amplification while transcriptionally limiting HR machinery expression during recombinant protein production to improve genetic stability.

7.2 Strain development for the increased production of secreted recombinant proteins

Whilst the secretion of recombinant proteins into extracellular media is an attractive option in the *Pichia* system to simplify downstream processing, redirecting large fluxes of proteins through the secretory pathway of *P. pastoris* introduces a number of process bottlenecks and product loss points that negatively affect final titres. The engineering of various aspects of the secretory pathway to adapt its capacity for larger volumes of protein has therefore formed a major focal point of *P. pastoris* research over the past decade and is ongoing with a number of currently active research projects within the field (Gasser, 2016; Puxbaum et al, 2016; Zahrl et al, 2016). Addressing the problems surrounding the trafficking of recombinant proteins through the ER formed one of the initial objectives within this project, since it is one of the most researched causes of secretion bottlenecks in

P. pastoris yet remains to be fully resolved. That being said, a strategy that has achieved past success in increasing recombinant protein secretion has entailed constitutively expressing ER resident UPR effectors, helping in protein folding and processing, that are normally only upregulated during ER stress (Gasser et al, 2005; Inan et al, 2006; Gasser et al, 2007B). In this vein a homologue belonging to the Opi1 transcription factor family was tested as a potential candidate to enhance the ER's capacity, as its deletion in *S. cerevisiae* uncoupled UPR-mediated ER expansion from the ER stress response and was shown to increase productivity. However the *Δopi1* phenotype was strongly concluded to be detrimental to recombinant protein secretion, instead causing increased bottlenecking in the secretory pathway and higher levels of UPR activity. The results indicate that the Opi1p in *P. pastoris* may regulate a different subset of genes governing cellular processes interacting with the secretory pathway to help maintain its native function. Alternatively it could be possible that the deletion of *OPI1* causes a remodeling of the ER but that, unlike *S. cerevisiae*, its new physiology grants it a reduced capacity for protein folding and processing, thereby worse adapting it to the high volumes of recombinant protein entering the secretory system during *AOX1*-driven expression.). One issue to note however is that all specific productivities and cellular growth values calculated in each study within this project used the optical density of cultures at 600nm (OD_{600}) as a measure for cell number. Whilst it provides a valid measure of cell density, changes in cell size that often occur under different growth conditions or as a direct result of the removal of certain genes will affect the OD_{600} of a cell culture irrespective of the actual cell count. Whilst this risk was minimised throughout the studies investigating clonal variation by maintaining consistent growth conditions between control and test samples, changes in the cell size of *Δopi1* mutants were not tested for and could therefore potentially have influenced the results if found to differ from the native strain. However the relationships between the specific productivities for the strains reported closely resembled the differences between their respective volumetric productivities, which provide a measure of total protein titre and do not account for cell density, supporting the validity of the observations presented in this study.

Throughout the course of this project it became increasingly clear that a major process flaw in the *Pichia* system was the use of expression media containing methanol as a sole carbon source for growth and the induction of transgene expression. While *P. pastoris* is classed as methylotrophic, methanol is far from the organism's most preferred carbon source and its metabolism via the Mut pathway creates a detrimental physiological background with knock-on effects on growth, viability, cellular stress and ultimately, productivity (Inan and Meagher, 2001; Vanz et al, 2012; Edwards-Jones

et al, 2015). This discovery has spawned a range of improved fermentation strategies centred around growing cells in mixed feeds of methanol and other carbon sources during expression (Zhang et al, 2003; Solà et al, 2007; Paulová et al, 2012; Niu et al, 2013). Even within this study a significant enhancement in protein production during batch expressions could be observed when methanol expression media was supplemented with either glucose or sorbitol. However considerable limitations are placed on the maximum efficacy of mixed feed strategies as growth on primary carbon sources results in the transcriptional repression of methanol-inducible promoters, including the popular *AOX1* promoter (Inan and Meagher, 2001). Based on a collection of recent studies mapping the regulatory network of P_{AOX1}, this study switched focus to alleviating glucose-induced repression of *AOX1* expression to generate new platform strains capable of growing on less limiting concentrations of glucose without compromising on protein production. The constitutive expression of a transcriptionally regulated global activator of Mut pathway genes – Mit1, resulted in a >20% increase in the specific productivity of cells grown in a mix of 0.5% (v/v) methanol and 0.5% (w/v) glucose, while the modifications made to a second global regulator – Mxr1p, show potential in reducing glucose repression further. A potential challenge arising from constitutively expressing its transcriptional enhancers is that the discrete repressed/induced states conferring the switch-like nature of P_{AOX1} could be compromised, as an increase in leaky expression on media lacking methanol was observed. However, the independent pathways in which P_{AOX1} is regulated by different carbon sources works to the benefit of this study and it is hypothesised that tight repression of P_{AOX1} in glucose repression resistant strains could be restored through redesigning the pre-culture medium to contain other repressive carbon sources. Examples include ethanol, which represses P_{AOX1} through inhibiting Mxr1p binding (Parua et al, 2012), and glycerol, which has been demonstrated to cause repression mediated by Mit1 (Wang et al, 2016).

Concepts for reducing glucose repression, or repression from other carbon sources, are unfortunately limited to the current understanding of the full regulatory network of P_{AOX1}, which is still incomplete. Transcriptomic profiles of *P. pastoris* grown in methanol have contributed to communal knowledge (Prielhofer et al, 2015) but many of the interactions that control the inhibition, activation and promoter binding of the global regulators of methanol inducible genes at the post-translational level remain unknown. Whilst the intracellular detection and transport of glucose plays a part in triggering glucose repression (Zhang et al, 2010; Polupanov, Nazarko, and Sibirny, 2012), the signalling pathways connecting carbon source sensing to transcription factor activation/ inhibition is yet to be elucidated. Little to no information is available on the 3D structures, activation domains, phosphorylated sites or

the existence of binding partners for the three main activators of P_{AOX1}: Mxr1p, Trm1p or Mit1. Research into each of these areas would not only give a better insight into how one of the most favoured promoters for recombinant expression in the *Pichia* system is modulated, but would also provide a host of new candidates and targets to engineer glucose repression out of P_{AOX1} regulation.

7.3 Final conclusions

Investigation into clonal variation within this project has observed that through seemingly stochastic events, the cause of which is independent to significant copy number variation, clones with an extreme deviation in productivity from the majority of the population arise at relatively high frequencies. These findings concur with previous research conducted by this group on the subject (Aw, 2012) and contribute by providing evidence to confirm that the use of zeocin as a selective agent is not a causative factor in clonal variation. Considerable variability in other native factors was also observed within clonal populations in this study but only differences in growth rate displayed any correlation to productivity, suggesting that differences in substrate metabolism between individual clones provide a small contributing factor amongst other traits to the variation in protein production between clones. The abnormally large differences between the highest and lowest producers found within the larger samples tested as part of this project show that caution must be taken when examining changes in the specific productivity of strains and that variability between biological replicates within samples needs to be critically assessed before making valid conclusions. Whilst it is difficult to recommend a representative sample size for a clonal population, practices such as controlling for transgene copy number, using more replicates, random selection as opposed to screening for the highest producers, anomaly detection and appropriate statistical testing are all vital in minimising the bias from clonal variation when evaluating strain improvement strategies for *P. pastoris*.

The transformation of any biological organism into a successful expression system requires the development of processes that grant high production consistencies, for which clonal variation presents an issue, but also the engineering of cells to prioritise the production of a foreign product over their own native functions. Considerable progress has been made in the repurposing of the secretory system of *P. pastoris* to export far larger quantities of proteins exceeding its own native requirements. Originally postulated as a method to further improve recombinant protein secretion, this project has now eliminated the deletion of the transcription factor *OPI1* as a strategy for increasing specific productivity. However the engineering of *P. pastoris* to reduced glucose repression of the *AOX1* promoter, presented in this study, can help inform the development of new strains that are capable of higher levels of methanol induced expression in mixed carbon source fermentations, which have fast become a new standard for protein production in the *Pichia* system.

8. References

Abad, S., Kitz, K., Hörmann, A., Schreiner, U., Hartner, F.S., and Glieder, A. (2010). Real-time PCR-based determination of gene copy numbers in *Pichia pastoris*. Biotechnol. J. *5*, 413–420.

Adrio, J.L., and Demain, A.L. (2014). Microbial enzymes: tools for biotechnological processes. Biomolecules *4*, 117–139.

Ahmad, M., Hirz, M., Pichler, H., and Schwab, H. (2014). Protein expression in *Pichia pastoris*: Recent achievements and perspectives for heterologous protein production. Appl. Microbiol. Biotechnol. *98*, 5301–5317.

Ambroziak, J., and Henry, S.A. (1994). INO2 and INO4 gene products, positive regulators of phospholipid biosynthesis in *Saccharomyces cerevisiae*, form a complex that binds to the INO1 promoter. J. Biol. Chem. *269*, 15344–15349.

Arico, C., Bonnet, C., and Javaud, C. (2013). N-glycosylation humanization for production of therapeutic recombinant glycoproteins in Saccharomyces cerevisiae. Methods Mol. Biol. *988*, 45–57.

Aw, R. (2012). Factors Affecting the Specific Productivity of *Pichia pastoris*. PhD thesis, Imperial College London

Aw, R., and Polizzi, K.M. (2013). Can too many copies spoil the broth? Microb. Cell Fact. *12*, 128.

Aw, R., and Polizzi, K.M. (2016). Liquid PTVA: a faster and cheaper alternative for generating multi-copy clones in *Pichia pastoris*. Microb. Cell Fact. *15*, 29.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Baumann, K., Adelantado, N., Lang, C., Mattanovich, D., and Ferrer, P. (2011). Protein trafficking , ergosterol biosynthesis and membrane physics impact recombinant protein secretion in *Pichia pastoris*.

Berdy, J. (1980) Bleomycin-Type Antibiotics. Amino Acid and Peptide Antibiotics, J. Berdy, ed. (Boca Raton, FL: CRC Press), 459-497.

Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th edition. New York: W H Freeman; 2002. Section 24.2, Amino Acids Are Made from Intermediates of the Citric Acid Cycle and Other Major Pathways. Available from: http://www.ncbi.nlm.nih.gov/books/NBK22459/

Biogrammatics. A short history of *Pichia pastoris*. Available at: http://www.biogrammatics.com/categories/*Pichia*-pastoris-history (accessed 21 August 2013)

Block-Alper, L., P. Webster, X. Zhou, L. Supekovà, W.-H. Wong, P.G. Schultz, and D.I. Meyer. (2002). IN02, a positive regulator of lipid biosynthesis, is essential for the formation of inducible membranes in yeast. Mol. Biol. Cell. *13*, 40–51.

Blount, B.A., Weenink, T., and Ellis, T. (2012). Construction of synthetic regulatory networks in yeast. FEBS Lett. *586*, 2112–2121.

Boehm, R. (2007). Bioproduction of therapeutic proteins in the 21st century and the role of plants and plant cells as production platforms. Ann. N. Y. Acad. Sci. *1102*, 121–134.

Brankamp, R.G., Sreekrishna, K., Smith, P.L., Blankenship, D.T., and Cardin, A.D. (1995). Expression of a Synthetic Gene Encoding the Anticoagulant-Antimetastatic Protein Ghilanten by the Methylotropic Yeast*Pichia pastoris*. Protein Expr. Purif. *6*, 813–820.

Bren, L. (2006). The road to the biotech revolution: highlights of 100 years of biologics regulation. FDA Consum. *40*, 50–57.

Brickner, J.H., and Walter, P. (2004). Gene recruitment of the activated INO1 locus to the nuclear membrane. PLoS Biol. *2*.

Buchetics, M., Dragosits, M., Maurer, M., Rebnegger, C., Porro, D., Sauer, M., Gasser, B., and Mattanovich, D. (2011). Reverse engineering of protein secretion by uncoupling of cell cycle phases from growth. Biotechnol. Bioeng. *108*, 2403–2412.

Bushell, M.E., Rowe, M., Avignone-Rossa, C.A., and Wardell, J.N. (2003). Cyclic fed-batch culture for production of human serum albumin in *Pichia pastoris*. Biotechnol. Bioeng. *82*, 678–683.

Camattari, A., Goh, A., Yip, L.Y., Tan, A.H.M., Ng, S.W., Tran, A., Liu, G., Liachko, I., Dunham, M.J., and Rancati, G. (2016). Characterization of a panARS-based episomal vector in the methylotrophic yeast *Pichia pastoris* for recombinant protein production and synthetic biology applications. Microb. Cell Fact. *15*, 139.

Celik, E., and Calık, P. (2012). Production of recombinant proteins by yeast cells. Biotechnol. Adv. *30*, 1108–1118.

Chen, R. (2012). Bacterial expression systems for recombinant protein production: *E. coli* and beyond. Biotechnol. Adv. *30*, 1102–1107.

Chen, Y., Cino, J., Hart, G., Freedman, D., White, C., and Komives, E.A. (1997). High protein expression in fermentation of recombinant *Pichia pastoris* by a fed-batch process. Process Biochem. *32*, 107–111.

Chen, Y.L., De Bernardis, F., Yu, S.J., Sandini, S., Kauffman, S., Tams, R.N., Bethea, E., and Reynolds, T.B. (2015). Candida albicans OPI1 regulates filamentous growth and virulence in vaginal infections, but not inositol biosynthesis. PLoS One *10*, 1–16.

Chen, C.C., Wu, P.H., Huang, C.T., and Cheng, K.J. (2004). A *Pichia pastoris* fermentation strategy for enhancing the heterologous expression of an Escherichia coli phytase. Enzyme Microb. Technol. *35*, 315–320.

Chi, Z., He, S., and Yao, S. (2005). Effects of *Pichia pastoris* INO1 expression in Schizosaccharomyces pombe on phosphatidylinositol (PI) synthesis and expression of INV+ encoding invertase. Enzyme Microb. Technol. *37*, 395–401.

Chiruvolu, V., Cregg, J.M., and Meagher, M.M. (1997). Recombinant protein production in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. Enzyme Microb. Technol. *21*, 277–283.

Chumnanpuen, P., Nookaew, I., and Nielsen, J. (2013). Integrated analysis, transcriptome-lipidome, reveals the effects of INO-level (INO2 and INO4) on lipid metabolism in yeast. BMC Syst. Biol. *7*, S7.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science (80-.). *339*, 819–823.

Cooper GM (2000). The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates. Recombination Between Homologous DNA Sequences. Available from: http://www.ncbi.nlm.nih.gov/books/NBK9859/

Cos, O., Serrano, A., Montesinos, J.L., Ferrer, P., Cregg, J.M., and Valero, F. (2005). Combined effect of the methanol utilization (Mut) phenotype and gene dosage on recombinant protein production in *Pichia pastoris* fed-batch cultures. J. Biotechnol. *117*, 321–335.

Couderc, R., and Baratti, J. (1980). Oxidation of Methanol by the Yeast, *Pichia pastoris*. Purification and Properties of the Alcohol Oxidase. Agric. Biol. Chem. *44*, 2279–2289.

Cox, J.S., Chapman, R.E., and Walter, P. (1997). The unfolded protein response coordinates the production of endoplasmic reticulum protein and endoplasmic reticulum membrane. Mol. Biol. Cell *8*, 1805–1814.

Cox, J.S., Shamu, C.E., and Walter, P. (1993). Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. Cell *73*, 1197–1206.

Cox, J.S., and Walter, P. (1996). A novel mechanism for regulating activity of a transcription factor that controls the unfolded protein response. Cell *87*, 391–404.

Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M., and Walter, P. (2005). On the mechanism of sensing unfolded protein in the endoplasmic reticulum. Proc. Natl. Acad. Sci. U. S. A. *102*, 18773–18784.

Cregg, J.M. (1986). Autonomous replication sequences for yeast strains of the genus *Pichia*. Patent EP 0180899 A2

Cregg, J.M., Barringer, K.J., Hessler, A.Y., and Madden, K.R. (1985). *Pichia pastoris* as a host system for transformations. Mol. Cell. Biol. 5, 3376–3385.

Cregg, J.M., Cereghino, J.L., Shi, J., and Higgins, D.R. (2000). Recombinant Protein Expression in *Pichia pastoris*. *16*.

Cregg, J.M. and Madden, K.R. (1987). Development of yeast transformation systems and construction of methanol-utilization-defective mutants of *Pichia pastoris* by gene disruption. Biological Research on Industrial Yeasts (Stewart, G.G., Russell, I., Klein, R.D. and Hiebsch, R.R., Eds.). *2*, 1–18.

Cregg, J.M., and Russell, K. a (2007). DNA mediated transformation. Methods Mol Biol. 2007, 389: pp 27-42. 10.1007/978-1-59745-456-8_3.

Clare, J.J., Rayment, F.B., Ballantine, S.P., Sreekrishna, K., and Romanos, M.A. (1991). High-level expression of tetanus toxin fragment C in *Pichia pastoris* strains containing multiple tandem integrations of the gene. Biotechnology. (N. Y). *9*, 455–460.

Critchlow, S.E., and Jackson, S.P. (1998). DNA-end-joining: from yeast to man. Trends Biochem. Sci. *23*, 394–398.

Daley, J.M., Palmbos, P.L., Wu, D., and Wilson, T.E. (2005). Nonhomologous end joining in yeast. Annu. Rev. Genet. *39*, 431–451.

Daly, R. and Hearn, M. T. W. (2005), Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. J. Mol. Recognit. *18*, 119–138.

Damasceno, L., Anderson, K., Ritter, G., Cregg, J., Old, L., and Batt, C. (2007). Cooverexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. Appl. Microbiol. Biotechnol. *74*, 381–389.

De Pourcq, K., De Schutter, K., and Callewaert, N. (2010). Engineering of glycosylation in yeast and other fungi: Current state and perspectives. Appl. Microbiol. Biotechnol. *87*, 1617–1631.

de Ruijter, J. C., Koskela, E.V. and Frey, A.D. (2016). Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. Microb. Cell Fact. *15:87*, 1–18.

De Schutter, K., Lin, Y.-C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouzé, P., Van de Peer, Y., and Callewaert, N. (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. Nat. Biotechnol. *27*, 561–566.

de Souza, P.M., and de Oliveira Magalhães, P. (2010). Application of microbial α-amylase in industry – A review. Brazilian J. Microbiol. *41*, 850–861.

Delic, M., Gongrich, R., Mattanovich, D., and Gasser, B. (2014). Engineering of protein folding and secretion-strategies to overcome bottlenecks for efficient production of recombinant proteins. Antioxid. Redox Signal. *21*, 414–437.

Deutscher, J. (2008). The mechanisms of carbon catabolite repression in bacteria. Curr. Opin. Microbiol. *11*, 87–93.

Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336–4343.

Dingwall, C., and Laskey, R.A. (1991). Nuclear targeting sequences — a consensus? Trends Biochem. Sci. *16*, 478–481.

Drocourt, D., Calmels, T., Reynes, J.P., Baron, M., and Tiraby, G. (1990). Cassettes of the Streptoalloteichus hindustanus ble gene for transformation of lower and higher eukaryotes to phleomycin resistance. Nucleic Acids Res. *18*, 4009.

Ecker, D.M., Jones, S.D., and Levine, H.L. (2015). The therapeutic monoclonal antibody market. MAbs *7*, 9–14.

Edwards-Jones, B., Aw, R., Barton, G.R., Tredwell, G.D., Bundy, J.G., and Leak, D.J. (2015). Translational arrest due to cytoplasmic redox stress delays adaptation to growth on methanol and heterologous protein expression in a typical fed-batch culture of *Pichia pastoris*. PLoS One *10*, 1–25.

Elsworth, R., Williams, V. and Harris-Smith, R. (1957), The effect of oxygen supply on the rate of growth of aerobacter cloacae. J. Appl. Chem *7*, 269–274.

Frand, A.R., and Kaiser, C.A. (1998). The ERO1 Gene of Yeast Is Required for Oxidation of Protein Dithiols in the Endoplasmic Reticulum. Mol. Cell *1*, 161–170.

Fredlund, E., Blank, L.M., Schnürer, J., Schnu, J., and Sauer, U. (2004). Oxygen- and Glucose-Dependent Regulation of Central Carbon Metabolism in *Pichia anomala*. Appl. Environ. Microbiol. *70*, 59'5–5911.

Fribley, A., Zhang, K., and Kaufman, R. (2009). Regulation of Apoptosis by the Unfolded Protein Response. In Apoptosis SE - 14, P. Erhardt, and A. Toth, eds. (Humana Press), pp. 191–204.

Friedlander, R., Jarosch, E., Urban, J., Volkwein, C., and Sommer, T. (2000). A regulatory link between ER-associated protein degradation and the unfolded-protein response. Nat. Cell Biol. *2*, 379–384.

Froger, A., and Hall, J.E. (2007). Transformation of Plasmid DNA into E . coli Using the Heat Shock Method. 2007.

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and *in vivo* for CRISPR-mediated genome editing. J. Integr. Plant Biol. *56*, 343–349.

Gaffney, P.J., and Curtis, A.D. (1987). A collaborative study to establish the 2nd international standard for tissue plasminogen activator (t-PA). Thromb. Haemost. *58*, 1085—1087.

Gatignol, A., Durand, H., and Tiraby, G. (1988). Bleomycin resistance conferred by a drug-binding protein. FEBS Lett. *230*, 171–175.

Gasser, B. (2016). Engineering protein quality control mechanisms and vesicular sorting pathways for efficient secretion of recombinant proteins. 14th International Congress on Yeasts. Awaji, Hyogo, Japan.

Gasser, B., Maurer, M., Gach, J., Kunert, R., Mattanovich, D. (2005). Engineering of *Pichia pastoris* for improved production of antibody fragments. Biotechnol. Bioeng., *94*: 353–361.

Gasser, B., Maurer, M., Rautio, J., Sauer, M., Bhattacharyya, A., Saloheimo, M., Penttilä, M., and Mattanovich, D. (2007) (A). Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. BMC Genomics *8*, 179.

Gasser, B., Sauer, M., Maurer, M., Stadlmayr, G., and Mattanovich, D. (2007) (B). Transcriptomics-based identification of novel factors enhancing heterologous protein secretion in yeasts. Appl. Environ. Microbiol. *73*, 6499–6507.

Gaughan, C.L. (2016). The present state of the art in expression , production and characterization of monoclonal antibodies. Mol. Divers. *20*, 255–270.

Geier, M., Fauland, P., Vogl, T., and Glieder, A. (2015). Compact multi-enzyme pathways in *P. pastoris*. Chem. Commun. *51*, 1643–1646.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth *6*, 343–345.

Gleeson, M.A.G., White, C.E., Meininger, D.P., and Komives, E.A. (1998). Generation of Protease-Deficient Strains and Their Use in Heterologous Protein Expression. In *Pichia* Protocols, D.R. Higgins, and J.M. Cregg, eds. (Totowa, NJ: Humana Press), pp. 81–94.

Goldstein, A.L., and McCusker, J.H. (1999). Three new dominant drug resistance cassettes for gene disruption in *Saccharomyces cerevisiae*. Yeast *15*, 1541–1553.

Grand View Research (2016). Industrial Enzymes Market Analysis By Product (Carbohydrase, Lipases, Proteases, Polymerases & Nudeases and Others), By Application (Textile, Feed Additive and Food Processing), By End-Use (Food & Beverage, Detergents, Animal Feed, Textile, Paper & Pulp, Nutraceutical, Personal Care & Cosmetics, and Wastewater) And Segment Forecasts To 2024. Available at http://www.grandviewresearch.com/industry-analysis/industrial-enzymes-market [Accessed 10 August 2016]

Grillitsch, K., Tarazona, P., Klug, L., Wriessnegger, T., Zellnig, G., Leitner, E., Feussner, I., and Daum, G. (2014). Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*. Biochim. Biophys. Acta *1838*, 1889–1897.

Guerfal, M., Ryckaert, S., Jacobs, P.P., Ameloot, P., Van Craenenbroeck, K., Derycke, R., and Callewaert, N. (2010). The HAC1 gene from *Pichia pastoris*: characterization and effect of its

overexpression on the production of secreted, surface displayed and membrane proteins. Microb. Cell Fact. *9*, 49.

Guffanti, E., Ferrari, R., Preti, M., Forloni, M., Harismendy, O., Lefebvre, O., and Dieci, G. (2006). A minimal promoter for TFIIIC-dependent in vitro transcription of snoRNA and tRNA genes by RNA polymerase III. J. Biol. Chem. *281*, 23945–23957.

Guo, Y., Lu, F., Zhao, H., Tang, Y., and Lu, Z. (2010). Cloning and heterologous expression of glucose oxidase gene from Aspergillus niger Z-25 in *Pichia pastoris*. Appl. Biochem. Biotechnol. *162*, 498–509.

Hampton, R.Y. (2002). ER-associated degradation in protein quality control and cellular regulation. Curr. Opin. Cell Biol. *14*, 476–482.

Hartner, F.S., and Glieder, A. (2006). Regulation of methanol utilisation pathway genes in yeasts. Microb. Cell Fact. *5*, 39.

Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., Lin-Cereghino, J., Kovar, K., Cregg, J.M., and Glieder, A. (2008). Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. Nucleic Acids Res. *36*, e76.

Hasslacher, M., Schall, M., Hayn, M., Bona, R., Rumbold, K., Lückl, J., Griengl, H., Kohlwein, S.D., and Schwab, H. (1997). High-Level Intracellular Expression of Hydroxynitrile Lyase from the Tropical Rubber TreeHevea brasiliensisin Microbial Hosts. Protein Expr. Purif. *11*, 61–71.

Hayworth, D. Overview of Protein Expression [online]. Available at http://www.piercenet.com/method/overview-protein-expression-systems [accessed 10 August, 2013]

Higgins, D.R., Busser, K., Comiskey, J., Whittier, P.S., Purcell, T.J. and Hoeffler, J.P. (1998) Small vectors for expression based on dominant drug resistance with direct multicopy selection. Methods Mol. Biol. 103, 41–53.

Higgins, D.R and Cregg, J.M eds. (1998) *Pichia* Protocols. Totowa, New Jersey: Humana Press.

Hirakawa, K., Kobayashi, S., Inoue, T., Endoh-Yamagami, S., Fukuda, R., and Ohta, A. (2009). Yas3p, an opi1 family transcription factor, regulates Cytochrome P450 expression in response to n-alkanes in Yarrowia lipolytica. J. Biol. Chem. *284*, 7126–7137.

Holst, J., Vignali, K.M., Burton, A.R., and Vignali, D.A.A. (2006). Rapid analysis of T-cell selection *in vivo* using T cell-receptor retrogenic mice. Nat Meth *3*, 191–197.

Idiris, A., Tohda, H., Bi, K., Isoai, A., Kumagai, H., and Giga-Hama, Y. (2006). Enhanced productivity of protease-sensitive heterologous proteins by disruption of multiple protease genes in the fission yeast Schizosaccharomyces pombe. Appl. Microbiol. Biotechnol. *73*, 404–420.

Inan, M. (2000). Studies on the alcohol oxidase (AOX1) promoter of *Pichia pastoris*. Dissertation. Lincoln, Nebraska: University of Nebraska. 101 pp.

Inan, M., Aryasomayajula, D., Sinha, J., and Meagher, M.M. (2006). Enhancement of protein secretion in *Pichia pastoris* by overexpression of protein disulfide isomerase. Biotechnol. Bioeng. *93*, 771–778.

Inan, M., and Meagher, M.M. (2001). Non-Repressing Carbon Sources for the Alcohol Oxidase (AOXI) Promoter of *Pichia pastoris*. Pap. Biochem. Eng. *92*, 585–589.

Ismail, N., and Ng, D.T.W. (2006). Have you HRD? Understanding ERAD Is DOAble! Cell *126*, 237–239.

Jahic, M., Gustavsson, M., Jansen, A.K., Martinelle, M., and Enfors, S.O. (2003). Analysis and control of proteolysis of a fusion protein in *Pichia pastoris* fed-batch processes. J. Biotechnol. *102*, 45–53.

Jahic, M., Veide, A., Charoenrat, T., Teeri, T., and Enfors, S.-O. (2006). Process technology for production and recovery of heterologous proteins with *Pichia pastoris*. Biotechnol. Prog. *22*, 1465–1473.

Jiang, W., Brueggeman, A.J., Horken, K.M., Plucinak, T.M., and Weeks, D.P. (2014). Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. Eukaryot. Cell *13*, 1465–1469.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Jusiak, B., Cleto, S., Perez-Piñera, P., and Lu, T.K. (2016). Engineering Synthetic Gene Circuits in Living Cells with CRISPR Technology. Trends Biotechnol. *34*, 535–547.

Kalderon, D., Richardson, W.D., Markham, A.F., and Smith, A.E. (1984). Sequence requirements for nuclear location of simian virus 40 large-T antigen. Nature 311, 33–38.

Kercher, M.A., Lu, P., and Lewis, M. (1997). Lac repressor-operator complex. Curr. Opin. Struct. Biol. *7*, 76–85.

Khanna, K.K., and Jackson, S.P. (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nat. Genet. *27*, 247–254.

Khow, O., and Suntrarachun, S. (2012). Strategies for production of active eukaryotic proteins in bacterial expression system. Asian Pac. J. Trop. Biomed. *2*, 159–162.

Kimura, M., Takatsuki, A., and Yamaguchi, I. (1994). Blasticidin S deaminase gene from *Aspergillus terreus* (BSD): a new drug resistance gene for transfection of mammalian cells. Biochim. Biophys. Acta *1219*, 653–659.

Kinch, M.S. (2015). An overview of FDA-approved biologics medicines. Drug Discov. Today *20*, 393–398.

Koning, A.J., Lum, P.Y., Williams, J.M., and Wright, R. (1993). $DiOC_6$ staining reveals organelle structure and dynamics in living yeast cells. Cell Motil. Cytoskeleton *25*, 111–128.

Kornberg, R.D. (1999). Eukaryotic transcriptional control. Trends Biochem. Sci. *24*, 46–49.

Kosugi S., Hasebe M., Tomita M., and Yanagawa H. (2009) Systematic identification of yeast cell cycle-dependent nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. USA *106*, 10171-10176.

Koutz, P., Davis, G.R., Stillman, C., Barringer, K., Cregg, J., and Thill, G. (1989). Structural comparison of the *Pichia pastoris* alcohol oxidase genes. Yeast *5*, 167–177.

Krainer, F.W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., and Glieder, A. (2012). Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. Microb. Cell Fact. *11*, 22.

Kranthi, B.V., Balasubramanian, N., and Rangarajan, P.N. (2006). Isolation of a single-stranded DNA-binding protein from the methylotrophic yeast, *Pichia pastoris* and its identification as zeta crystallin. Nucleic Acids Res. *34*, 4060–4068.

Kranthi, B.V., Kumar, R., Kumar, N.V., Rao, D.N., and Rangarajan, P.N. (2009). Identification of key DNA elements involved in promoter recognition by Mxr1p, a master regulator of methanol utilization pathway in *Pichia pastoris*. Biochim. Biophys. Acta - Gene Regul. Mech. *1789*, 460–468.

Kuberl, A., Schneider, J., Thallinger, G.G., Anderl, I., Wibberg, D., Hajek, T., Jaenicke, S., Brinkrolf, K., Goesmann, A., Szczepanowski, R., et al. (2011). High-quality genome sequence of *Pichia pastoris* CBS7435. J. Biotechnol. *154*, 312–320.

Kumar, N.V., and Rangarajan, P.N. (2012). The zinc finger proteins Mxr1p and repressor of phosphoenolpyruvate carboxykinase (ROP) have the same DNA binding specificity but regulate methanol metabolism antagonistically in *Pichia pastoris*. J. Biol. Chem. *287*, 34465–34473.

Kumita, J.R., Johnson, R.J.K., Alcocer, M.J.C., Dumoulin, M., Holmqvist, F., McCammon, M.G., Robinson, C. V., Archer, D.B., and Dobson, C.M. (2006). Impact of the native-state stability of human lysozyme variants on protein secretion by *Pichia pastoris*. FEBS J. *273*, 711–720.

Kurjan, J., and Herskowitz, I. (1982). Structure of a yeast pheromone gene (MF alpha): a putative alpha-factor precursor contains four tandem copies of mature alpha-factor. Cell *30*, 933–943.

Kurtzman, C.P. (2005). Description of Komagataella phaffii sp. nov. and the transfer of *Pichia* pseudopastoris to the methylotrophic yeast genus Komagataella. Int. J. Syst. Evol. Microbiol. *55*, 973–976.

Kurtzman, C.P. (2009). Biotechnological strains of Komagataella (*Pichia*) pastoris are Komagataella phaffii as determined from multigene sequence analysis. J. Ind. Microbiol. Biotechnol. *36*, 1435–1438.

Ladisch, M.R., and Kohlmann, K.L. (1992). Recombinant human insulin. Biotechnol. Prog. *8*, 469–478.

Larsen, S., Weaver, J., de Sa Campos, K., Bulahan, R., Nguyen, J., Grove, H., Huang, A., Low, L., Tran, N., Gomez, S., et al. (2013). Mutant strains of *Pichia pastoris* with enhanced secretion of recombinant proteins. Biotechnol. Lett. *35*, 10.1007/s10529–013 – 1290–1297.

Larson, M.H., Gilbert, L. a, Wang, X., Lim, W. a, Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. *8*, 2180–2196.

Laukens, B., De Wachter, C., and Callewaert, N. (2015). Engineering the *Pichia pastoris* N-Glycosylation Pathway Using the GlycoSwitch Technology. In Glyco-Engineering: Methods and Protocols, A. Castilho, ed. (New York, NY: Springer New York), pp. 103–122.

Lee, S.Y. (1996). High cell-density culture of Escherichia coli. Trends Biotechnol. *14*, 98–105.

Lee, Y.C., and Yang, D. (2002). Determination of lysozyme activities in a microplate format. Anal. Biochem. *310*, 223–224.

Levene, H. (1960). Robust tests for equality of variances. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press. 278–292.

Liachko, I., Youngblood, R.A., Tsui, K., Bubb, K.L., Queitsch, C., Raghuraman, M.K., Nislow, C., Brewer, B.J., and Dunham, M.J. (2014). GC-Rich DNA Elements Enable Replication Origin Activity in the Methylotrophic Yeast *Pichia pastoris*. PLoS Genet. *10*.

Lin-Cereghino, G.P., Godfrey, L., de la Cruz, B.J., Johnson, S., Khuongsathiene, S., Tolstorukov, I., Yan, M., Lin-Cereghino, J., Veenhuis, M., Subramani, S., et al. (2006). Mxr1p, a key regulator of the methanol utilization pathway and peroxisomal genes in *Pichia pastoris*. Mol. Cell. Biol. *26*, 883–897.

Lin-Cereghino, G.P., Lin-Cereghino, J., Ilgen, C., and Cregg, J.M. (2002). Production of recombinant proteins in fermenter cultures of the yeast *Pichia pastoris*. Curr Opin Biotech *13*, 329–332.

Lin-Cereghino, G.P., Lin Cereghino, J., Jay Sunga, A., Johnson, M. a., Lim, M., Gleeson, M. a G., and Cregg, J.M. (2001). New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of *Pichia pastoris*. Gene 263, 159–169.

Lodish H, Berk A, Zipursky SL, et al. (2000) Molecular Cell Biology. 4th edition. New York: W. H. Freeman;. Section 17.3, Overview of the Secretory Pathway. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21471/

Loewen, C.J.R. (2004). Phospholipid Metabolism Regulated by a Transcription Factor Sensing Phosphatidic Acid. Science (80-.). *304*, 1644–1647.

Lõoke, M., Kristjuhan, K., and Kristjuhan, A. (2011). Extraction of genomic DNA from yeasts for PCR-based applications. Biotechniques *50*, 325–328.

Love, K.R., Panagiotou, V., Jiang, B., Stadheim, T. a, and Love, J.C. (2010). Integrated single-cell analysis shows *Pichia pastoris* secretes protein stochastically. Biotechnol. Bioeng. *106*, 319–325.

Love, K.R., Politano, T.J., Panagiotou, V., Jiang, B., Stadheim, T. a, and Love, J.C. (2012). Systematic single-cell analysis of *Pichia pastoris* reveals secretory capacity limits productivity. PLoS One *7*, e37915.

Ma, Q., Teter, B., Ubeda, O.J., Morihara, T., Dhoot, D., Nyby, M.D., Tuck, M.L., Frautschy, S.A., and Cole, G.M. (2009). Direct Selection of *Pichia pastoris* expression strains using new G418 resistance vectors. NIH Public Access. 27, 14299–14307.

Ma, H., Wu, Y., Dang, Y., Choi, J.-G., Zhang, J., and Wu, H. (2014). Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. Mol. Ther. Nucleic Acids *3*, e161.

Macauley-Patrick, S., Fazenda, M.L., McNeil, B., and Harvey, L.M. (2005). Heterologous protein production using the *Pichia pastoris* expression system. Yeast *22*, 249–270.

Maccani, A., Landes, N., Stadlmayr, G., Maresch, D., Leitner, C., Maurer, M., Gasser, B., Ernst, W., Kunert, R., and Mattanovich, D. (2014). *Pichia pastoris* secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins. Biotechnol. J. *9*, 526–537.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-Guided Human Genome Engineering via Cas9. Science. *339*, 823–826.

Marfori, M., Mynott, A., Ellis, J.J., Mehdi, A.M., Saunders, N.F.W., Curmi, P.M., Forwood, J.K., Bod??n, M., and Kobe, B. (2011). Molecular basis for specificity of nuclear import and prediction of nuclear localization. Biochim. Biophys. Acta - Mol. Cell Res. *1813*, 1562–1577.

Marx, H., Mecklenbräuker, A., Gasser, B., Sauer, M., and Mattanovich, D. (2009). Directed gene copy number amplification in *Pichia pastoris* by vector integration into the ribosomal DNA locus. FEMS Yeast Res. *9*, 1260–1270.

Malherbe, D.F., du Toit, M., Cordero Otero, R.R., van Rensburg, P., and Pretorius, I.S. (2003). Expression of the Aspergillus niger glucose oxidase gene in *Saccharomyces cerevisiae* and its potential applications in wine production. Appl. Microbiol. Biotechnol. *61*, 502–511.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.

Mattanovich, D., Callewaert, N., Rouzé, P., Lin, Y.-C., Graf, A., Redl, A., Tiels, P., Gasser, B., and De Schutter, K. (2009). Open access to sequence: browsing the *Pichia pastoris* genome. Microb. Cell Fact. *8*, 53.

Mattanovich, D., Gasser, B., Hohenblum, H., and Sauer, M. (2004). Stress in recombinant protein producing yeasts. J. Biotechnol. *113*, 121–135.

Meyer, H., and Schmidhalter, D.R. (2012). Microbial Expression Systems and Manufacturing from a Market and Economic Perspective, Innovations in Biotechnology, Dr. Eddy C. Agbo (Ed.), ISBN: 978-953-51- 0096-6, InTech, Available from: http://www.intechopen.com/books/innovations-in-biotechnology/microbial- expression-systems-and-manufacturing-from-a-market-and-economic-perspective

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology *155*, 733–740.

Mokdad-Gargouri, R., Abdelmoula-Soussi, S., Hadiji-Abbes, N., Amor, I.Y.-H., Borchani-Chabchoub, I., and Gargouri, A. (2012). Yeasts as a tool for heterologous gene expression. Methods Mol. Biol. *824*, 359–370.

Näätsaari, L., Mistlberger, B., Ruth, C., Hajek, T., Hartner, F.S., and Glieder, A. (2012). Deletion of the *Pichia pastoris* KU70 homologue facilitates platform strain generation for gene expression and synthetic biology. PLoS One *7*, e39720.

Nakamura, Y., Nishi, T., Noguchi, R., Ito, Y., Watanabe, T., Nishiyama, T., Aikawa, S., Hasunuma, T., Ishii, J., Okubo, Y., and Kondo, A. (2016). A new autonomous replicating plasmid vector containing centromere DNA sequence of *Pichia pastoris*. 14th International Congress on Yeasts. Awaji, Hyogo, Japan.

Nasser, M.W., Pooja, V., Abdin, M.Z., and Jain, S.K. (2003). Evaluation of Yeast as an Expression System. *2*, 477–493.

National Institute for Health and Care Excellence (2015). NICE: Kadcyla price still too high. Available at https://www.nice.org.uk/news/press-and-media/nice-kadcyla-price-still-too-high [Accessed 10 August 2016]

Nett, J.H., Cook, W.J., Chen, M.T., Davidson, R.C., Bobrowicz, P., Kett, W., Brevnova, E., Potgieter, T.I., Mellon, M.T., and Prinz, B. (2013). Characterization of the *Pichia pastoris* Protein-O-mannosyltransferase Gene Family. PLoS One 8, 1–13.

Niu, H., Jost, L., Pirlot, N., Sassi, H., Daukandt, M., Rodriguez, C., and Fickers, P. (2013). A quantitative study of methanol/sorbitol co-feeding process of a *Pichia pastoris* Mut+/pAOX1-lacZ strain. Microb. Cell Fact. *12*, 33.

Norden, K., Agemark, M., Danielson, J.A.H., Alexandersson, E., Kjellbom, P., and Johanson, U. (2011). Increasing gene dosage greatly enhances recombinant expression of aquaporins in *Pichia pastoris*. BMC Biotechnol. *11*, 47.

Normington, K., Kohno, K., Kozutsumi, Y., Gething, M.J., and Sambrook, J. (1989). *S. cerevisiae* encodes an essential protein homologous in sequence and function to mammalian BiP. Cell *57*, 1223–1236.

Oliva-Trastoy, M., Trastoy, M.O., Defais, M., and Larminat, F. (2005). Resistance to the antibiotic Zeocin by stable expression of the Sh ble gene does not fully suppress Zeocin-induced DNA cleavage in human cells. Mutagenesis *20*, 111–114.

Otto, R., Santagostino, A., and Schrader, U. (2014). Rapid growth in biopharma: Challenges and opportunities. McKinsey&Company. Available at http://www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/rapid-growth-in-biopharma [Accessed 10 August 2016]

Parua, P.K., Ryan, P.M., Trang, K., and Young, E.T. (2012). *Pichia pastoris* 14-3-3 regulates transcriptional activity of the methanol inducible transcription factor Mxr1 by direct interaction. Mol. Microbiol. *85*, 282–298.

Patil, C., and Walter, P. (2001). Intracellular signaling from the endoplasmic reticulum to the nucleus: the unfolded protein response in yeast and mammals. Curr. Opin. Cell Biol. *13*, 349–355.

Paulová, L., Hyka, P., Branská, B., Melzoch, K., and Kovar, K. (2012). Use of a mixture of glucose and methanol as substrates for the production of recombinant trypsinogen in continuous cultures with *Pichia pastoris* Mut +. J. Biotechnol. *157*, 180–188.

Persistence Market Research, (2015). Global Biopharmaceuticals Market Will Reach US$ 278 Bn by 2020. Available at http://www.persistencemarketresearch.com/mediarelease/biopharmaceutical-market.asp [Accessed 10 August 2016]

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. *29*, e45.

Pincus, D., Chevalier, M.W., Aragón, T., van Anken, E., Vidal, S.E., El-Samad, H., and Walter, P. (2010). BiP binding to the ER-stress sensor Ire1 tunes the homeostatic behavior of the unfolded protein response. PLoS Biol. *8*, e1000415.

Pla, I.A., Damasceno, L.M., Vannelli, T., Ritter, G., Batt, C.A., and Shuler, M.L. (2006). Evaluation of Mut+ and MutS *Pichia pastoris* phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration. Biotechnol. Prog. *22*, 881–888.

Polupanov, A.S., Nazarko, V.Y., and Sibirny, A.A. (2012). Gss1 protein of the methylotrophic yeast *Pichia pastoris* is involved in glucose sensing, pexophagy and catabolite repression. Int. J. Biochem. Cell Biol. *44*, 1906–1918.

Pope, B., and Kent, H.M. (1996). High efficiency 5 min transformation of Escherichia coli. Nucleic Acids Res. *24*, 536–537.

Prielhofer, R., Cartwright, S.P., Graf, A.B., Valli, M., Bill, R.M., Mattanovich, D., and Gasser, B. (2015). *Pichia pastoris* regulates its gene-specific response to different carbon sources at the transcriptional, rather than the translational, level. BMC Genomics *16*, 167.

Protein expression in *Pichia* (2013). Research Corporation Technologies [online]. Available at: http://www.*Pichia*.com/science-center/ (accessed 01 August 2016).

Puxbaum, V., Gasser, B., and Mattanovich, D. (2016). The bud tip is the cellular hot spot of protein secretion in yeasts. Appl. Microbiol. Biotechnol. *100*, 8159–8168.

Puxbaum, V., Landes, N., Gasser, B., and Mattanovich, D. (2016). A winding road to the bud tip – exploring the secretion route of recombinant proteins in *Pichia pastoris* via live cell microscopy. 14th International Congress on Yeasts. Awaji, Hyogo, Japan.

Puxbaum, V., Mattanovich, D., and Gasser, B. (2015). Quo vadis? The challenges of recombinant protein folding and secretion in *Pichia pastoris*. Appl. Microbiol. Biotechnol. *99*, 2925–2938.

Qin, X., Qian, J., Yao, G., Zhuang, Y., Zhang, S., and Chu, J. (2011). GAP promoter library for fine-tuning of gene expression in *Pichia pastoris*. Appl. Environ. Microbiol. *77*, 3600–3608.

Radcliffe, P.A., and Mitrophanous, K.A. (2004). Multiple gene products from a single vector: "self-cleaving" 2A peptides. Gene Ther. *11*, 1673–1674.

Rahdar, M., McMahon, M.A., Prakash, T.P., Swayze, E.E., Bennett, C.F., and Cleveland, D.W. (2015). Synthetic CRISPR RNA-Cas9-guided genome editing in human cells. Proc. Natl. Acad. Sci. *112*, E7110–E7117.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rebnegger, C., Graf, A.B., Valli, M., Steiger, M.G., Gasser, B., Maurer, M., and Mattanovich, D. (2014). In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. Biotechnol. J. *9*, 511–525.

Romanos, M., Scorer, C., Sreekrishna, K., and Clare, J. (1998). The Generation of Multicopy Recombinant Strains. In *Pichia* Protocols, D.R. Higgins, and J.M. Cregg, eds. (Totowa, NJ: Humana Press), pp. 55–72.

Rosano, G.L., and Ceccarelli, E.A. (2014). Recombinant protein expression in Escherichia coli : Advances and challenges. Front. Microbiol. *5*, 1–17.

Rossanese, O.W., Soderholm, J., Bevis, B.J., Sears, I.B., O'Connor, J., Williamson, E.K., and Glick, B.S. (1999). Golgi structure correlates with transitional endoplasmic reticulum organization in *Pichia pastoris* and *Saccharomyces cerevisiae*. J. Cell Biol. *145*, 69–81.

Ruth, C., Buchetics, M., Vidimce, V., Kotz, D., Naschberger, S., Mattanovich, D., Pichler, H., and Gasser, B. (2014). *Pichia pastoris* Aft1 - a novel transcription factor, enhancing recombinant protein secretion. Microb. Cell Fact. *13*, 120.

Ryan, M.D., King, A.M.Q., and Thomas, G.P. (1991). Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence. J. Gen. Virol. *72*, 2727–2732.

Sabnis, R.W., Deligeorgiev, T.G., Jachak, M.N., and Dalvi, T.S. (1997). DiOC6(3): a Useful Dye for Staining the Endoplasmic Reticulum. Biotech. Histochem. *72*, 253–258.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotech *32*, 347–355.

Sasano, Y., Yurimoto, H., Yanaka, M., and Sakai, Y. (2008). Trm1p, a Zn(II)2Cys6-type transcription factor, is a master regulator of methanol-specific gene activation in the methylotrophic yeast Candida boidinii. Eukaryot. Cell *7*, 527–536.

Sahu, U., Krishna Rao, K., and Rangarajan, P.N. (2014). Trm1p, a Zn(II)2Cys6-type transcription factor, is essential for the transcriptional activation of genes of methanol utilization pathway, in *Pichia pastoris*. Biochem. Biophys. Res. Commun. *451*, 158–164.

Schenk, J., Balazs, K., Jungo, C., Urfer, J., Wegmann, C., Zocchi, A., Marison, I. W. and von Stockar, U. (2008), Influence of specific growth rate on specific productivity and glycosylation of a recombinant avidin produced by a *Pichia pastoris* Mut+ strain. Biotechnol. Bioeng. *99*, 368–377

Schmidt, F.R. (2004). Recombinant expression systems in the pharmaceutical industry. Appl. Microbiol. Biotechnol. *65*, 363–372.

Schuck, S., Prinz, W. a, Thorn, K.S., Voss, C., and Walter, P. (2009). Membrane expansion alleviates endoplasmic reticulum stress independently of the unfolded protein response. J. Cell Biol. *187*, 525–536.

Schwarzhans, J.-P., Wibberg, D., Winkler, A., Luttermann, T., Kalinowski, J., and Friehs, K. (2016). Integration event induced changes in recombinant protein productivity in *Pichia pastoris* discovered by whole genome sequencing and derived vector optimization. Microb. Cell Fact. *15*, 84.

Scorer, C.A., Clare, J.J., McCombie, W.R., Romanos, M.A., and Sreekrishna, K. (1994). Rapid selection using G418 of high copy number transformants of *Pichia pastoris* for high-level foreign gene expression. Biotechnology. (N. Y). *12*, 181–184.

Sears, I.B., O'Connor, J., Rossanese, O.W., and Glick, B.S. (1998). A versatile set of vectors for constitutive and regulated gene expression in *Pichia pastoris*. Yeast *14*, 783–790.

Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). Escherichia coli physiology in Luria-Bertani broth. J. Bacteriol. *189*, 8746–8749.

Shaffer, A.L., Shapiro-Shelef, M., Iwakoshi, N.N., Lee, A.H., Qian, S.B., Zhao, H., Yu, X., Yang, L., Tan, B.K., Rosenwald, A., et al. (2004). XBP1, downstream of Blimp-1, expands the secretory apparatus and other organelles, and increases protein synthesis in plasma cell differentiation. Immunity *21*, 81–93.

Shamu, C.E., and Walter, P. (1996). Oligomerization and phosphorylation of the Ire1p kinase during intracellular signaling from the endoplasmic reticulum to the nucleus. EMBO J *15*, 3028-3039.

Silva, C.I.F., Teles, H., Moers, A.P.H.A., Eggink, G., De Wolf, F.A., and Werten, M.W.T. (2011). Secreted production of collagen-inspired gel-forming polymers with high thermal stability in *Pichia pastoris*. Biotechnol. Bioeng. *108*, 2517–2525.

Simonen, M., and Palva, I. (1993). Protein secretion in *Bacillus* species. Microbiol. Mol. Biol. Rev. *57*, 109–137.

Sinha, J., Plantz, B.A., Inan, M., and Meagher, M.M. (2005). Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon-τ Biotechnol. Bioeng. *89*, 102–112.

Solà, A., Jouhten, P., Maaheimo, H., Sánchez-Ferrando, F., Szyperski, T., and Ferrer, P. (2007). Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. Microbiology *153*, 281–290.

Sonoda, E., Hochegger, H., Saberi, A., Taniguchi, Y., and Takeda, S. (2006). Differential usage of non-homologous end-joining and homologous recombination in double strand break repair. DNA Repair (Amst). *5*, 1021–1029.

Stadlmayr, G., Mecklenbräuker, A., Rothmüller, M., Maurer, M., Sauer, M., Mattanovich, D., and Gasser, B. (2010). Identification and characterisation of novel *Pichia pastoris* promoters for heterologous protein production. J. Biotechnol. *150*, 519–529.

Stasyk, O. V., Stasyk, O.G., Komduur, J., Veenhuis, M., Cregg, J.M., and Sibirny, A.A. (2004). A Hexose Transporter Homologue Controls Glucose Repression in the Methylotrophic Yeast Hansenula polymorpha. J. Biol. Chem. *279*, 8116–8125.

Steinborn, G., Böer, E., Scholz, A., Tag, K., Kunze, G., and Gellissen, G. (2006). Application of a wide-range yeast vector (CoMed) system to recombinant protein production in dimorphic Arxula adeninivorans, methylotrophic Hansenula polymorpha and other yeasts. Microb. Cell Fact. *5*, 33.

Stotz, A., and Linder, P. (1990). The ADE2 gene from *Saccharomyces cerevisiae*: sequence and new vectors. Gene *95*, 91–98.

Sturmberger, L., Chappell, T., Geier, M., Krainer, F., Day, K.J., Vide, U., Trstenjak, S., Schiefer, A., Richardson, T., Soriaga, L., et al. (2016). Refined *Pichia pastoris* reference genome sequence. J. Biotechnol. 1–11.

Sukhacheva, M. V., Davydova, M.E., and Netrusov, a. I. (2004). Production of Penicillium funiculosum 433 Glucose Oxidase and Its Properties. Appl. Biochem. Microbiol. *40*, 25–29.

Sunga, A.J., Tolstorukov, I., and Cregg, J.M. (2008). Posttransformational vector amplification in the yeast *Pichia pastoris*. FEMS Yeast Res. *8*, 870–876.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A.A. (2004). Correction of multi-gene deficiency *in vivo* using a single "self-cleaving" 2A peptide-based retroviral vector. Nat. Biotechnol. *22*, 589–594.

Takagi, S., Tsutsumi, N., Terui, Y., and Kong, X.Y. (2012). Method for methanol independent induction from methanol inducible promoters in *Pichia*. Patent US 8236528 B2

Thor, D., Xiong, S., Orazem, C.C., Kwan, A.C., Cregg, J.M., Lin-Cereghino, J., and Lin-Cereghino, G.P. (2005). Cloning and characterization of the *Pichia pastoris* MET2 gene as a selectable marker. FEMS Yeast Res. *5*, 935–942.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249–258.

Tschopp, J.F., Brust, P.F., Cregg, J.M., Stillman, C.A., and Gingeras, T.R. (1987). Expression of the lacZ gene from two methanol-regulated promoters in *Pichia pastoris*. Nucleic Acids Res. *15*, 3859–3876.

U.S. Food and Drug Administration, 2015. Novel Drug Approvals for 2015. Available at http://www.fda.gov/Drugs/DevelopmentApprovalProcess/DrugInnovation/ucm430302.htm [Accessed 10 August 2016)

Valero, F. (2012). Heterologous expression systems for lipases: a review. Methods Mol. Biol. *861*, 161–178.

Vanz, A.L., Lünsdorf, H., Adnan, A., Nimtz, M., Gurramkonda, C., Khanna, N., and Rinas, U. (2012). Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. Microb. Cell Fact. *11*, 103.

Vervecken, W., Callewaert, N., Kaigorodov, V., Geysens, S., and Contreras, R. (2007). Modification of the N-Glycosylation Pathway to Produce Homogeneous, Human-Like Glycans Using GlycoSwitch Plasmids. In *Pichia* Protocols, J.M. Cregg, ed. (Totowa, NJ: Humana Press), pp. 119–138.

Viader-Salvadó, J.M., Cab-Barrera, E.L., Galán-Wong, L.J., and Guerrero-Olazarán, M. (2006). Genotyping of recombinant *Pichia pastoris* strains. Cell. Mol. Biol. Lett. *11*, 348–359.

Vink, T., Oudshoorn-Dickmann, M., Roza, M., Reitsma, J.-J., and de Jong, R.N. (2013). A simple, robust and highly efficient transient expression system for producing antibodies. Methods.

Vogl, T., and Glieder, A. (2013). Regulation of *Pichia pastoris* promoters and its consequences for protein production. N. Biotechnol. *30*, 385–404.

Vyas, V.K., Barrasa, M.I., and Fink, G.R. (2015). A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families. Sci. Adv. *1*, e1500248.

Wagner, C., and Dietz, M. (2001). The negative regulator Opi1 of phospholipid biosynthesis in yeast contacts the pleiotropic repressor Sin3 and the transcriptional activator Ino2. *41*, 155–166.

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. Nat. Biotechnol. *32*, 992–1000.

Wan, H., Sjölinder, M., Schairer, H.U., and Leclerque, A. (2004). A new dominant selection marker for transformation of *Pichia pastoris* to soraphen A resistance. J. Microbiol. Methods 57, 33–39.

Wang, X., Cai, M., Shi, L., Wang, Q., Zhu, J., Wang, J., Zhou, M., Zhou, X., and Zhang, Y. (2015). PpNrg1 is a transcriptional repressor for glucose and glycerol repression of AOX1 promoter in methylotrophic yeast *Pichia*pastoris. Biotechnol. Lett.

Wang, X., Wang, Q., Wang, J., Bai, P., Shi, L., Shen, W., Zhou, M., Zhou, X., Zhang, Y., and Cai, M. (2016). Mit1 Transcription Factor Mediates Methanol Signaling and Regulates *Alcohol Oxidase 1* Promoter in *Pichia pastoris*. J. Biol. Chem. jbc.M115.692053.

Weninger, A., Glieder, A., and Vogl, T. (2015) (A). A toolbox of endogenous and heterologous nuclear localization sequences for the methylotrophic yeast *Pichia pastoris*. FEMS Yeast Res. *15*, fov082.

Weninger, A., Hatzl, A.M., Schmid, C., Vogl, T., and Glieder, A. (2015) (B). Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast *Pichia pastoris*. J. Biotechnol.

Westers, L., Westers, H., and Quax, W.J. (2004). Bacillus subtilis as a cell factory for pharmaceutical proteins: A biotechnological approach to optimize the host organism. Biochim. Biophys. Acta - Mol. Cell Res. *1694*, 299–310.

White, M.J., Hirsch, J.P., and Henry, S.A. (1991). The OPI1 gene of *Saccharomyces cerevisiae*, a negative regulator of phospholipid biosynthesis, encodes a protein containing polyglutamine tracts and a leucine zipper. J. Biol. Chem. *266*, 863–872.

Whyteside, G., Alcocer, M.J.C., Kumita, J.R., Dobson, C.M., Lazarou, M., Pleass, R.J., and Archer, D.B. (2011). Native-state stability determines the extent of degradation relative to secretion of protein variants from *Pichia pastoris*. PLoS One *6*, e22692.

Wiest, D.L., Burkhardt, J.K., Hester, S., Hortsch, M., Meyer, D.I., and Argon, Y. (1990). Membrane biogenesis during B cell differentiation: Most endoplasmic reticulum proteins are expressed coordinately. J. Cell Biol. *110*, 1501–1511.

Wilkinson, B., and Gilbert, H.F. (2004). Protein disulfide isomerase. BBA Proteins and Proteomics. *1699*, 35–44.

Wriessnegger, T., Gübitz, G., Leitner, E., Ingolic, E., Cregg, J., de la Cruz, B.J., and Daum, G. (2007). Lipid composition of peroxisomes from the yeast *Pichia pastoris* grown on different carbon sources. Biochim. Biophys. Acta - Mol. Cell Biol. Lipids *1771*, 455–461.

Wu, S., and Letchworth, G.J. (2004). High efficiency transformation by electroporation of *Pichia pastoris* pretreated with lithium acetate and dithiothreitol. Biotechniques *36*, 152–154.

Wu, M., Shen, Q., Yang, Y., Zhang, S., Qu, W., Chen, J., Sun, H., and Chen, S. (2013). Disruption of YPS1 and PEP4 genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115. J. Ind. Microbiol. Biotechnol. *40*, 589–599.

Xiao, R., Wilkinson, B., Solovyov, A., Winther, J.R., Holmgren, A., Lundström-Ljung, J., and Gilbert, H.F. (2004). The contributions of protein bisulfide isomerase and its homologues to oxidative protein folding in the yeast endoplasmic reticulum. J. Biol. Chem. *279*, 49780–49786.

Xie, J., Zhou, Q., Du, P., Gan, R., and Ye, Q. (2005). Use of different carbon sources in cultivation of recombinant *Pichia pastoris* for angiostatin production. Enzyme Microb. Technol. *36*, 210–216.

Xuan, Y., Zhou, X., Zhang, W., Zhang, X., Song, Z., and Zhang, Y. (2009). An upstream activation sequence controls the expression of AOX1 gene in *Pichia pastoris*. FEMS Yeast Res. *9*, 1271–1282.

Yamada, Y., Maeda, K., and Mikata, K. (1994). The phylogenetic relationships of the hat-shaped ascospore-forming, nitrate-assimilating *Pichia* species, formerly classified in the genus Hansenula Sydow et Sydow, based on the partial sequences of 18S and 26S ribosomal RNAs (Saccharomycetaceae): the proposals of three new genera, Ogataea, Kuraishia, and Nakazawaea. Biosci. Biotechnol. Biochem. *58*, 1245–1257.

Yamada, Y., Matsuda, M., Maeda, K., and Mikata, K. (1995). The phylogenetic relationships of methanol-assimilating yeasts based on the partial sequences of 18S and 26S ribosomal RNAs: the proposal of Komagataella gen. nov. (Saccharomycetaceae). Biosci. Biotechnol. Biochem. *59*, 439–444.

Yan,M., Rachubinski, D. A., Joshi, S.,Rachubinski, R. A. and Subramani S.(2008).Dysferlin domain-containing proteins, Pex30p and Pex31p, localized to two compartments, control the number and size of oleate-induced peroxisomes in *Pichia pastoris*. Mol. Biol. Cell. *19*, 885–898.

Yang, J., Nie, L., Chen, B., Liu, Y., Kong, Y., Wang, H., and Diao, L. (2014). Hygromycin-resistance vectors for gene expression in *Pichia pastoris*. Yeast 31, 115–125.

Yin, J., Li, G., Ren, X., and Herrler, G. (2007). Select what you need : A comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *127*, 335–347.

Yip, C.L., Welch, S.K., Klebl, F., Gilbert, T., Seidel, P., Grant, F.J., O'Hara, P.J., and MacKay, V.L (1994). Cloning and analysis of the Saccharomyces cerevisiae MNN9 and MNN1 genes required for complex glycosylation of secreted proteins. Proc. Natl. Acad. Sci. U. S. A. *91*, 2723–2727.

Young, T.S., Ahmad, I., Brock, A., and Schultz, P.G. (2009). Expanding the genetic repertoire of the methylotrophic yeast *Pichia pastoris*. Biochemistry *48*, 2643–2653.

Yurimoto, H., Oku, M., and Sakai, Y. (2011). Yeast methylotrophy: Metabolism, gene regulation and peroxisome homeostasis. Int. J. Microbiol. *2011*.

Zahrl, R., Pfeffer, M., Mattanovich, D., and Gasser, B. (2016). The impact of ERAD on recombinant protein secretion in yeast. 14th International Congress on Yeasts. Awaji, Hyogo, Japan.

Zhang, W., Hywood Potter, K.J., Plantz, B. a, Schlegel, V.L., Smith, L. a, and Meagher, M.M. (2003). *Pichia pastoris* fermentation with mixed-feeds of glycerol and methanol: growth kinetics and production improvement. J. Ind. Microbiol. Biotechnol. *30*, 210–215.

Zhang, L., Jia, R., Palange, N.J., Satheka, A.C., Togo, J., An, Y., Humphrey, M., Ban, L., Ji, Y., Jin, H., et al. (2015). Large genomic fragment deletions and insertions in mouse using CRISPR/Cas9. PLoS One *10*, 1–14.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., Qiu, J.-L., and Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nat. Commun. *7*, 12617.

Zhang, P., Zhang, W., Zhou, X., Bai, P., Cregg, J.M., and Zhang, Y. (2010). Catabolite Repression of Aox in *Pichia pastoris* is dependent on hexose transporter PpHxt1 and pexophagy. Appl. Environ. Microbiol. *76*, 6108–6118.

Zhiyi, H., Wenshan, L., Wenze, Z., Ning, D., Chi, Z., Kenan, Y., Ping, W., Qianqian, W., and Qing, Z. (2007). Secretion expression and activity assay of a novel fusion protein of thrombopoietin and interleukin-6 in *Pichia pastoris*. J. Biochem. *142*, 17–24.

Zhu, T., Guo, M., Tang, Z., Zhang, M., Zhuang, Y., Chu, J., and Zhang, S. (2009) (A). Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast *Pichia pastoris*. J. Appl. Microbiol. *107*, 954–963.

Zhu, T., Guo, M., Sun, C., Qian, J., Zhuang, Y., Chu, J., and Zhang, S. (2009) (B). A systematical investigation on the genetic stability of multi-copy *Pichia pastoris* strains. Biotechnol. Lett. *31*, 679–684.

Zhu, T., Guo, M., Zhuang, Y., Chu, J., and Zhang, S. (2011). Understanding the effect of foreign gene dosage on the physiology of *Pichia pastoris* by transcriptional analysis of key genes. Appl. Microbiol. Biotechnol. *89*, 1127–1135.

Zinser, E., Sperka-gottlieb, C.D.M., Fasch, E., Kohlwein, S.D., Paltauf, F., Daum, G., Graz, T.U., and Graz, A.- (1991). Phospholipid Synthesis and Lipid Composition of Subcellular Membranes in the Unicellular Eukaryote *Saccharomyces cerevisiae*. *173*, 2026–2034.

Zito, E. (2015). ERO1: A protein disulfide oxidase and H_2O_2 producer. Free Radic. Biol. Med. *83*, 299–304.

9. Appendix

9.1 DNA vectors and oligomers used within this study

Table A1: Oligonucleotides used within this study. Lowercase sequence represents regions without homology to the intended template. All oligonucleotides were synthesised and purchased from Eurofins Genomics (Germany).

Name	5' - 3' Sequence	T_M (region of homology) /°C	Description
03-AmpR-F	CAAACAAACCACCGCTGG	65.7	forward primer - AmpR (including promoter and terminator region) from pUC18
04-AmpR-R	TTCAAATATGTATCCGCTCATG	61.7	reverse primer - AmpR (including promoter and terminator region) from pUC18
06-GibAmpRpαb-F	tgggacgctcgaaggctttaatttgcaagctggagaccaacatgtTTCAAATATGTATCCGCTCATG	61.7	forward Gibson primer for AmpR from pUC18
07-GibAmpRpαb-R	ccttttttacggttcctggccttttgctggccttttgctcacatgtCAAACAAACCACCGCTGGTA	66.1	reverse Gibson primer for AmpR from pUC18
08-GibGOx-F	agaaggggtatctctcgagaaaagaggctgaagctgcaATGGTGTCTGTATTTCTCAGCA	61.6	Gibson forward primer for the *GOX* gene from *P. funiculosum* to insert into the MCS of pPICZα
09-GibGOx-R	ggtaccgatccgagacggccggctgggccacgtgaattccCTAGGCACTTTTGGCATAGTC	60.8	Gibson reverse primer for the *GOX* gene from *P. funiculosum* to insert into the MCS of pPICZα
10-Gibkan-F	atatcggcatagtataatacgacaaggtgaggaactaaaccATGAGCCATATTCAACGGGA	64.5	Forward primer for k*anr* for Gibson assembly into pAmpZ
11-Gibkan-R	cgacaaaggaaaaggggacggatctccgaggcctgggacccgtgggccgccgtcggacgtgTTAGAAAAACTCATCGAGCATC	60.4	Reverse primer for *kanr* for Gibson assembly into pAmpZ
12-5'AOX1	GACTGGTTCCAATTGACAAGC	63.8	Forward primer for sequencing of vectors containing the *AOX1* promoter
13-3'AOX1	GCAAATGGCATTCTGACATCC	66.8	Reverse primer for sequencing of vectors containing the *AOX1* transcription terminator
14-Gibhis-F	atatcggcatagtataatacgacaaggtgaggaactaaaccATGACATTTCCCTTGCTACC	60.9	Forward primer for *HIS4* for Gibson assembly into pAmpZ

15-Gibhis-R	cgacaaaggaaaaggggacggatctccgaggcctgggacccgtgggccgccgtcggacgtgTTAAATAAGTCCCAGTTTCTCCA	60.8	Reverse primer for *HIS4* for Gibson assembly into pAmpZ
16-Selseqpb-F	GGCGGTGTTGACAATTAATC	61	Forward sequencing primer for genes within the pPICZαb/pAVE522's selective marker cassette
17-Selseqpb-R	TTACATGATATCGACAAAGG	56	Reverse primer for genes within the pPICZαb/pAVE522's selective marker cassette
18-Opi1kan 5'-F	TCGTACACGTCGCCAAACTTAT	65.6	Forward primer for 5' flanking region to *OPI1* in *P. pastoris* to assemble the *OPI1* knockout construct
19-Opi1kan 5'-R	aggacaccgctagcggatccGGACGCTGCTAGATAGAATTATGAG	63.3	Gibson reverse primer for 5' flanking region to *OPI1* in *P. pastoris* to assemble the *OPI1* knockout construct
20-Opi1kan-F	aattctatctagcagcgtccGCTAGCGGTGTCCTCGTC	64.2	Gibson forward primer for *Tn903kanr* as part of the *OPI1* knockout construct
21-Opi1kan-R	tccaattatttaatccaaatGAATTCAGGAGTGGGAAATACCA	65.2	Gibson reverse primer for *Tn903kanr* as part of the *OPI1* knockout construct
22-Opi1kan 3'-F	tatttcccactcctgaattcGGATTAAATAATTGGAAAAGTAGTAGTATAGGTAACG	65.1	Gibson forward primer for the 3' region flanking *OPI1* in *P. pastoris* to assemble *OPI1* knockout construct
23-Opi1kan 3'-R	AACGGATATTACTGCGGAGCC	66.3	Reverse primer for the 3' region flanking *OPI1* in *P. pastoris* to assemble *OPI1* knockout construct
24-GibAOXzeo-F	tggcccagccggccgtcagatcaaaaaacaactaattattcgaaaaaATGGCCAAGTTGACCAGTGC	67.7	Gibson forward primer for *sh ble* for assembly into the *AOX1* site of pAVE522K
25-GibAOXzeo-R	tagcggccgccgcggctcgtggtaccaggctacaaacagcTCAGTCCTGCTCCTCGGC	67.4	Gibson reverse primer for *sh ble* for assembly into the *AOX1* site of pAVE522K
26-Opi1-F	ATGAGTAAACGAAAACTTACAGGGG	64.3	Forward primer annealing to *OPI1* to test for its presence by PCR
27-Opi1-R	CTATTCTTGCTTTTCCCCCG	64.8	Reverse primer annealing to *OPI1* to test for its presence by PCR
28-Opi15'flankseq	ACCAAATTATCCCCCTCTATCGAAC	56	Forward primer annealing 50bp 5' to *OPI1* to amplify part of the *OPI1* locus
29-Opi13'flankseq	GAGGGAAAGACTGAAAGCTTTACAG	56	Reverse primer annealing 92bp 3' to *OPI1* to amplify part of the *OPI1* locus
30-qACT1-F	GCTTTGTTCCACCCATCTGT	56.9	qPCR forward primer for *ACT1*, producing a 163 bp amplicon. Efficiency = 91.05%

31-qACT1-R	TGCATACGCTCAGCAATACC	56.8	qPCR reverse primer for *ACT1*, producing a 163 bp amplicon. Efficiency = 91.05%
32-qHAC1-F	CGACTACATTACTACAGCTCCATCA	64	qPCR forward primer for *HAC1*, producing a 124bp amplicon. Efficiency = 89.24%
33-qHAC1-R	TGCTGTAATGTGTAAAGATGAATCC	61	qPCR reverse primer for *HAC1*, producing a 124bp amplicon. Efficiency = 89.24%
34-qKAR2-F	TCAAAGACGCTGGTGTCAAG	58	qPCR forward primer for *KAR2*, producing a 151bp amplicon. Efficiency = 89.04%
35-qKAR2-R	TATGCGACAGCTTCATCTGG	58	qPCR reverse primer for *KAR2*, producing a 151bp amplicon. Efficiency = 89.04%
36-qPDI-F	GCCGTTAAATTCGGTAAGCA	56	qPCR forward primer for *PDI*, producing a 145bp amplicon. Efficiency = 89.84%
37-qPDI-R	TCAGCTCGGTCACATCTTTG	58	qPCR reverse primer for *PDI*, producing a 145bp amplicon. Efficiency = 89.84%
40-qGOx2-F	CCGCACGAACAATATCAAGG	61.4	qPCR forward primer for *GOX*, producing a 120bp amplicon. Efficiency = 91.35%
41-qGOx2-R	TCCATGCCAAAGACCTTCTC	60.2	qPCR reverse primer for *GOX*, producing a 120bp amplicon. Efficiency = 91.35%
48-Gibhph-F	atatcggcatagtataatacgacaaggtgaggaactaaaccATGGGTAAAAAGCCTGAACTCAC	64.4	Gibson forward primer for *hph* for assembly into pPCas9
49-Gibhph-R	cgacaaaggaaaaggggacggatctccgaggcctgggacccgtgggccgccgtcggacgtgTTATTCCTTTGCCCTCGGAC	65.3	Gibson reverse primer for *hph* for Gibson assembly into pPCas9
50-GibGapfrag-F	GAGAGCTTCTTCTACGGCCC	64.6	Forward primer amplifying a 272bp 3' fragment of the *GAP* promoter for Gibson assembly into pPCas9
51-GibGapfrag-R	cttcttgtccatgtttaaacATAGTTGTTCAATTGATTGAAATAGGG	63.2	Reverse primer amplifying a 272bp 3' fragment of the *GAP* promoter for Gibson assembly into pPCas9
52-GibCas9-F	tcaatcaattgaacaactatgtttaaacATGGACAAGAAGTACTCCATTG	60	Gibson forward primer for *CAS9-SV40* with an MssI site for Gibson assembly downstream of P_{GAP} in pPCas9
53-GibCas9-R	gtaagtgcccaacttgaactgaggaacagtcatgtctagcggccgccgcggctcgtggtaccaggctacaaacgtttaaactCACACCTTCCTCTTCTTCTTG	61	Gibson reverse primer for *CAS9-SV40* with an MssI site for Gibson assembly downstream of P_{GAP} in pPCas9

55-Cas95'1kb-F	ATGGACAAGAAGTACTCCATTGG	63.3	RT-PCR forward primer to confirm the transcription of *CAS9*
56-Cas95'1kb-R	AAAGTCAAGTCTTGGTGGTGC	63.5	RT-PCR reverse primer to confirm the transcription of *CAS9*
57-sc-gRNA-F	TCTTTGAAAAGATAATGTATGATTA	56	PCR forward primer to amplify gRNA with the *S. cerevisiae SNR52* promoter and *SUP4* terminator
58-sc-gRNA-R	GGAAGTGAATGGAGACATAA	57.1	PCR reverse primer to amplify gRNA with the *S. cerevisiae SNR52* promoter and *SUP4* terminator
59-ribo-gRNA-F	TTTTTGTAGAAATGTCTTGGTGTC	61.2	PCR forward primer to amplify gRNA with the G1 promoter, ribozyme sites and *ADH* terminator
60-ribo-gRNA-R	TTGTCCTCTGAGGACATAAAATAC	60.4	PCR reverse primer to amplify gRNA with the G1 promoter, ribozyme sites and *ADH* terminator
61-ade2-F	ATGGATTCTCAGGTAATAGGTATTC	58.1	PCR forward primer to amplify the *P. pastoris ADE2* gene
62-ade2-R	TCAAAGACGATTCTTCAAATAG	59.7	PCR reverse primer to amplify the *P. pastoris ADE2* gene
63-ade2seq-F	GTTGAAGTCCCGGACTATG	54.1	Forward sequencing primer for *ADE2*
64-ade2seq-R	GCGTTGGTGTTGGAAGTAG	55.2	Reverse sequencing primer for *ADE2*
71-Gibmxr1p-F	tgggacgctcgaaggctttaatttgcaagctggagaccaaTGATCGGAACCCTGACC	62.8	Gibson forward primer for the assembly of the region -700 - +1200 of *MXR1,* with a 3' *SV40* signal, into pAVECRS
72-Gibmxr1sv40-R	atcataagaaattcgctcacaccttcctcttcttcttgggGCATGATAACGTGTTAGAGAAAGTC	62.1	Gibson reverse primer for the assembly of the region -700 - +1200 of *MXR1,* with a 3' *SV40* signal, into pAVECRS
73-Gibadh1tt-F	taacacgttatcatgccccaagaagaagaggaaggtgtgaGCGAATTTCTTATGATTTATGA	57.9	Gibson forward primer for the assembly of the *ADH1* transcription terminator downstream of *mxr1t-sv40* in pAVECRS
74-Gibadh1tt-R	tttacggttcctggccttttgctggccttttgctcacatgTTGTCCTCTGAGGACATAAA	57.8	Gibson reverse primer for the assembly of the *ADH1* transcription terminator downstream of *mxr1t-sv40* in pAVECRS
75-mxr1sv40seq-F	CTGTACAGACGCGTGTACG	56.9	Sequencing forward primer for *mxr1t-sv40* within pAVECRS
76-mxr1sv40seq-R	AGCGACCTCATGCTATACC	55.5	Sequencing reverse primer for *mxr1t-sv40* within pAVECRS
83-qsynHuL3-F	CGGGTACCGTGGTATCTCTTT	60.2	qPCR forward primer for *HUL T70A*, producing a 97bp amplicon. Efficiency = 99.26%
84-qsynHuL3-R	CCTGTCACCAGCGTTGTAGTT	60.2	qPCR reverse primer for *HUL T70A*,

			producing a 97bp amplicon. Efficiency = 99.26%
93-qshble2-F	AGTTGACCAGTGCCGTTCC	61.1	qPCR forward primer for *sh ble*, producing a 105bp amplicon. Efficiency = 93.02%
94-qshble2-R	AGTCGTCCTCCACGAAGTCC	61.6	qPCR reverse primer for *sh ble*, producing a 105bp amplicon. Efficiency = 93.02%
99-GibGapmit1-F	tttgtccctatttcaatcaattgaacaactatgtttaaacATGAGTACCGCAGCCC	60.4	Gibson forward primer for the assembly of *MIT1* downstream of P_{GAP} in pPCas9
100-Gibmit1AOX-R	gccgccgcggctcgtggtaccaggctacaaacgtttCTATTCTTCAACATTCCAGTAGTCA	60.5	Gibson reverse primer for the assembly of *MIT1* upstream of the *AOX1* transcription terminator in pPCas9
107-Gibmit1T2A	gaacccttccctcagcacgTTCTTCAACATTCCAGTAGTCAAT	60.9	Gibson reverse primer for the assembly of *MIT1* into the polycistronic expression cassette with *mxr1t-sv40*
108-GibT2A	ACTACTGGAATGTTGAAGAACGTGCTGAGGGAAGGGGTTCACTTCTAACCTGCGGTGACGTGGAGGAGAACCCCGGACCTATGAGCAATCTACCCCCAAC	N/A	Gibson part containing the *T. asigna* T2A sequence to assemble the polycistronic *MIT1/mxr1t-sv40* expression cassette
109-GibT2Amxr1t-F	tggaggagaaccccggacctATGAGCAATCTACCCCCAAC	63.1	Gibson forward primer for the assembly of *mxr1t-sv40* downstream of the T2A sequence within the polycistronic expression cassette
110-Gibmxr1tAOX1-R	gtctagcggccgccgcggctcgtggtaccaggctacaaacgtttaaacTCAGCATGATAACGTGTTAGAGAAAGTC	62.1	Gibson forward primer for the assembly of *mxr1t-sv40* upstream of the *AOX1* transcription terminator sequence within the polycistronic expression cassette
111-5'GAP	GTCCCTATTTCAATCAATTGAA	59.8	Forward primer for sequencing of vectors containing the *GAP* promoter
114-qhygb2-F	GACAATGGCCGCATAACAG	60.1	qPCR forward primer for *hph*, producing a 114bp amplicon. Efficiency = 89.18%
115-qhygb2-R	CTGCTCCATACAAGCCAACC	60.7	qPCR reverse primer for *hph*, producing a 114bp amplicon. Efficiency = 89.18%
116-qmit1-1-F	CCAAGCAGATCTGTGGGATT	60.1	qPCR forward primer for *MIT1*, producing a 193bp amplicon. Efficiency = 97.22%
117-qmit1-1-R	AACTCCTTGCCCTTCCAGTT	60.1	qPCR reverse primer for *MIT1*, 193bp amplicon. Efficiency = 97.22%

Table A2: DNA vectors/plasmids used within this study.

Plasmid name	Description	Source	Reference
pPICZαB	*P. pastoris* secreted expression vector utilising the *AOX1* promoter for targeted integration into the *AOX1* locus *sh ble;* pUC ori*;* 5'*AOX1-αmf-*MCS-*AOX1 TT* poly-histidine tag, c-myc epitope tag	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
pAmpZα	Modified from pPICZαB with an AmpR expression cassette cloned between the Sh ble expression cassette and pUC ori	This study	Unpublished
pGOxZα	Based on pAmpZα, with *gox* cloned into the MCS downstream of the AOX1 promoter	This study	Unpublished
pGOxHα	Modified from pAmpZα, with *sh ble* replaced with *HIS4* in the same expression cassette and *gox* cloned into the MCS downstream of the AOX1 promoter	This study	Unpublished
pAVE522	*P. pastoris* expression vector utilizing the AOX1 promoter for targeted integration into the AOX1 locus *sh ble;* bacterial ori*;* 5'*AOX1-*MCS-*AOX1 TT*	Fujifilm Diosynth Biotechnologies (Billingham, UK)	Unpublished
pAVE522-k	Modified from pAVE522, with *sh ble* replaced with *Tn903kan*^{r} in the same expression cassette	This study	Unpublished
pAVECRS	Modified from pAVE522-k, with *sh ble* cloned into the MCS, downstream of the *AOX1* promoter	This study	Unpublished

Plasmid name	Description	Source	Reference
pAVECRS-mxr1t-sv40	Modified from pAVECRS to contain an expression cassette consisting of an expression cassette for the N terminal 400 amino acids of Mxr1p with a C-terminal SV40 signal *P_{mxr1}-1200bp 5' mxr1-SV40-ADH TT*	This study	Unpublished
pIB2	*P. pastoris* expression vector utilizing the *GAP* promoter and integrating into the *HIS4* locus *HIS4*; pMB1 ori; *ampr*, 5'*GAP*-MCS-*AOX1 TT*	Gift from Benjamin Glick (Addgene plasmid #25451)	Sears et al. (1998)
pGrzαHSA	*P. pastoris* expression vector integrating into the rDNA locus, based on pPICZα. 5'AOX1 upstream region replaced with rDNA locus sequence and the *GAP* promoter from pIB2. *5' GAP-αmf-HSA-AOX1 TT*	Rochelle Aw, Imperial College	Aw (2012)
pAG32	Vector containing an *hph* expression cassette for *S. cerevisiae*, based on pFA6. *ampr*; pBR322 ori; *hph*	Gift from John McCusker (Addgene plasmid #35122)	Goldstein and McCusker (1999)
p414-TEF1p-Cas9-CYC1t	*S. cerevisiae* expression vector for human-optimised *S. pyogenes* Cas9. *ampr*; pBR322 ori, *5' TEF1-CAS9-sv40 NLS-CYC1 TT*; *TRP1*	Gift from George Church (Addgene plasmid # 43802)	DiCarlo et al (2013)

Plasmid name	Description	Source	Reference
pPCas9	*P. pastoris* expression vector for human-optimised *S. pyogenes* Cas9, modified from pGrzαHSA. *sh ble* replaced with *hph* and *αmf-HSA* replaced with *CAS9-sv40 NLS*	This study	Unpublished
pPhGmit1	Modified from pPCas9. *CAS9-SV40 NLS* replaced with *MIT1*	This study	Unpublished
pPhGmm1	Modified from pPCas9. *CAS9-SV40 NLS* replaced with *MIT1-T2A-MXR1* (5' 1200bp)	This study	Unpublished
pJET 1.2/blunt	*E. coli* blunt cloning vector with a lethal insert for positive selection of transformants cloned with blunt PCR products *ampr*; *eco47IR*; *pMB1* ori	Thermo Fisher Scientific	Thermo Fisher Scientific catalogue
pJET-snr52p-340-sup4t	Expression cassette consisting of the *S. cerevisiae snr52* promoter, gRNA targeting position 340 of *P. pastoris ade2* and *S. cerevisiae sup4* TT cloned into pJET 1.2/blunt	This study/ Gene Art, Thermo Fisher (UK)	Unpublished
pJET-snr52p-689-sup4t	Expression cassette consisting of the *S. cerevisiae snr52* promoter, gRNA targeting position 340 of *P. pastoris ade2* and *S. cerevisiae sup4* TT cloned into pJET 1.2/blunt	This study/ Gene Art, Thermo Fisher (UK)	Unpublished
pG1-340-adht	Expression cassette consisting of the *P. pastoris* G1 promoter - Hammerhead ribozyme - gRNA targeting position 340 of *P. pastoris ade2* - HDV ribozyme - *adh* TT cloned into pMK-RQ. *kanr*; Col E1 ori;	This study/ Gene Art, Thermo Fisher Scientific (UK)	Unpublished
pG1-689-adht	Expression cassette consisting of the *P. pastoris* G1 promoter - Hammerhead ribozyme - gRNA targeting position 689 of *P. pastoris ade2* - HDV ribozyme - *adh* TT cloned into pMK-RQ. *kanr*; Col E1 ori;	This study/ Gene Art, Thermo Fisher Scientific (UK)	Unpublished
pPICzm-T70A	*P. pastoris* expression vector based on pPICZα. pPICZα expression cassette replaced with pPIC9k-based expression cassette containing *HULT70A* downstream of the AOX1 promoter. 5' *AOX1-αmf-HULT70A-AOX1* TT	Bryn-Edwards Jones, Imperial College	Unpublished

9.2 Glucose oxidase activity assay standard curve

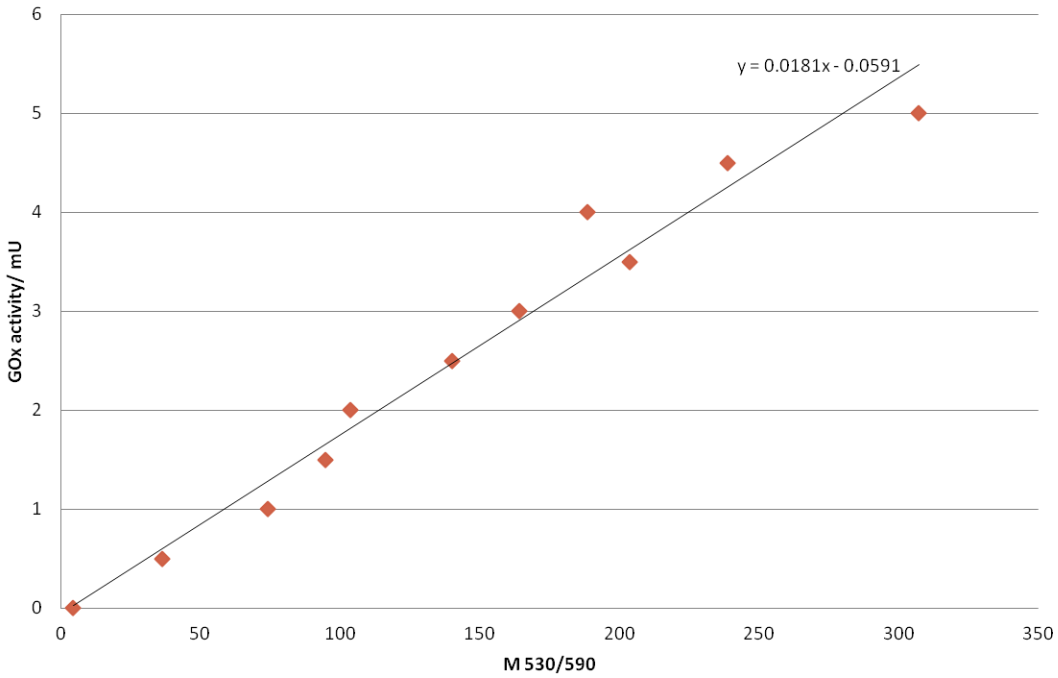


Figure 61: Standard curve plotting the emission values at 590nm after excitement at 530nm against the activity of known glucose oxidase standards from the Amplex Red Glucose oxidase assay kit (Thermo Fisher Scientific, UK).

9.3 Log-transformed growth curve of NRRL 11430 *Δopi1* on sorbitol

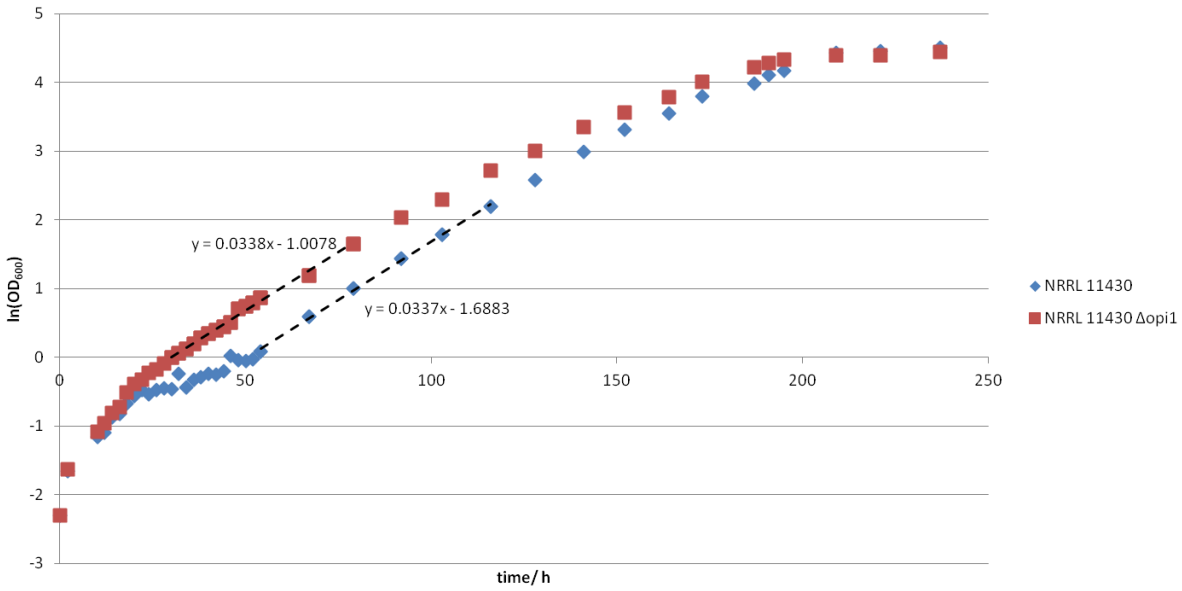


Figure A62: Growth curve displaying the natural logarithm (ln) of cell density for *P. pastoris* NRRL 11430 (WT) and NRRL 11430 *Δopi*1 against time during growth in minimal media containing 4% (w/v) sorbitol. The dotted line for each plot represents the range of data points used to calculate the doubling times of each strain.

9.4 Biological parts designed and synthesised for CRISPR-Cas9 of *P. pastoris*

9.4.1 *P. pastoris ADE2* gRNA target sequence locations

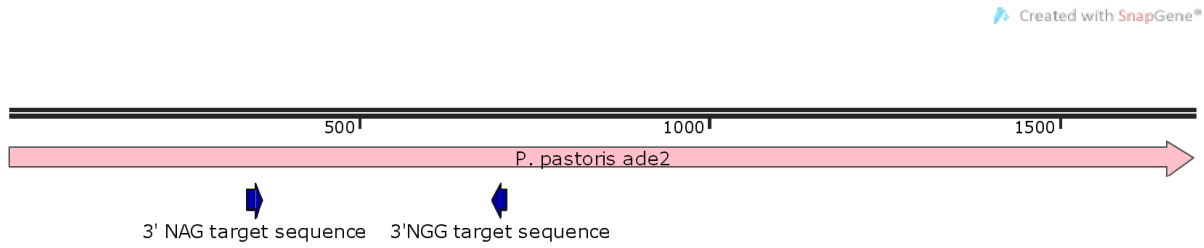


Figure A63: Map of *P. pastoris ADE2* showing the location of the 2 potential CRISPR recognition sequences used in this study. A 5'(N_{20})-NAG-3' site occurs at position 340 on the sense strand while a 5'(N_{20})-NGG-3' site was identified at position 689 on the antisense strand.

9.4.2 *ADE2* HDR template with a premature stop codon

caaggacaatatctgtcatattgtgtatgctccggccagagttaatgacaccatctgacaaaagaaagctcaaatattagctgaaaacactgtgaa

gactttcccaggcgctg

9.4.3 gRNA expression cassette design

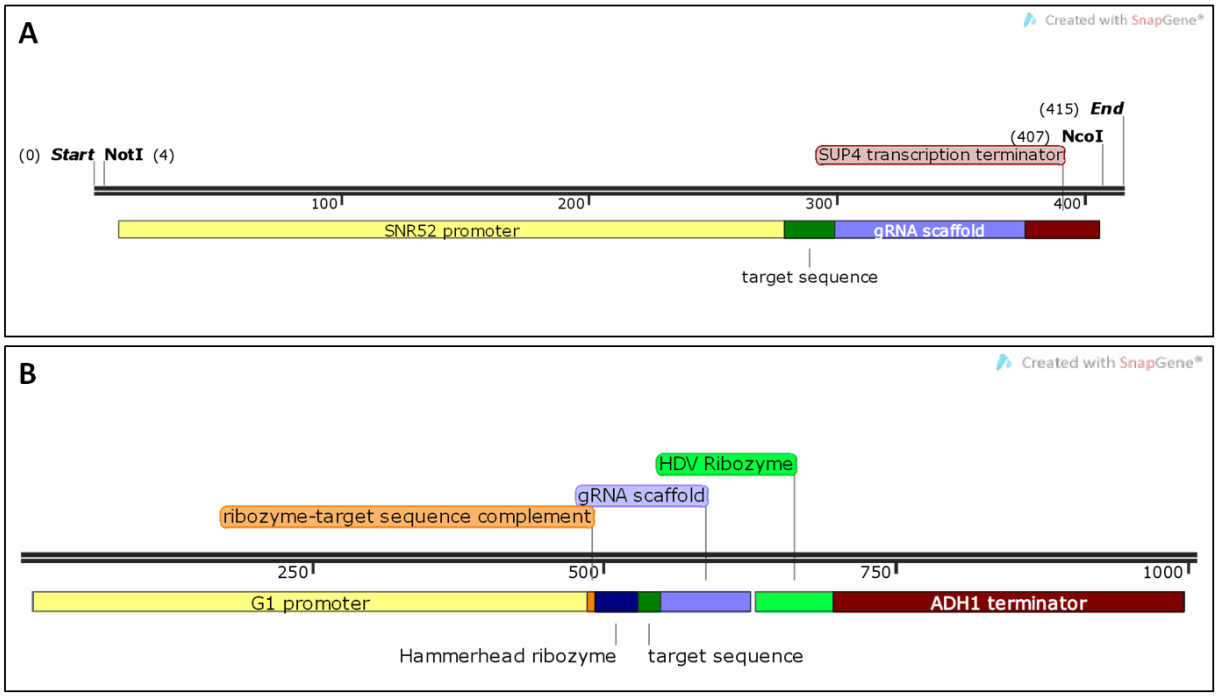


Figure A64: Generalised maps for expression cassettes employing the 2 different strategies for gRNA expression in *P. pastoris*. A – Expression cassette described by DiCarlo et al. (2013) for gRNA expression in *S. cerevisiae*, employing the yeast *SNR52* promoter for RNA polymerase III mediated transcription, and the *SUP4* transcription terminator. B – Expression cassette based on the system described by Gao and Zhao (2014) for expression using RNA polymerase II promoters. Mature gRNA sequences are flanked by self-processing ribozymes that catalyze their cleavage and removal from the resulting mRNA transcript.

9.4.4 Ribozyme mechanism of action

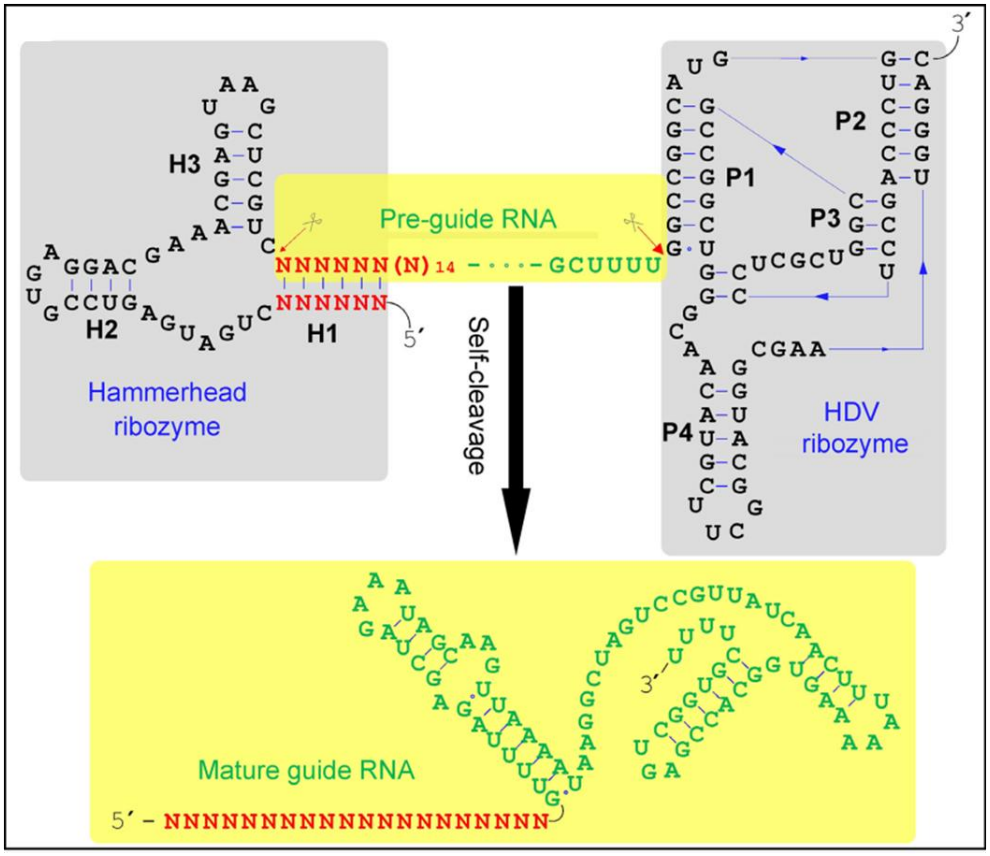


Figure A65: Schematic from Gao and Zhao (2014) illustrating the mechanism of self-cleavage of pre-guide mRNA transcripts flanked by the Hammerhead and HDV ribozymes to release the final gRNA transcript free of post-transcriptional modifications.

9.5 Classic nuclear localisation signal search in Mxr1p

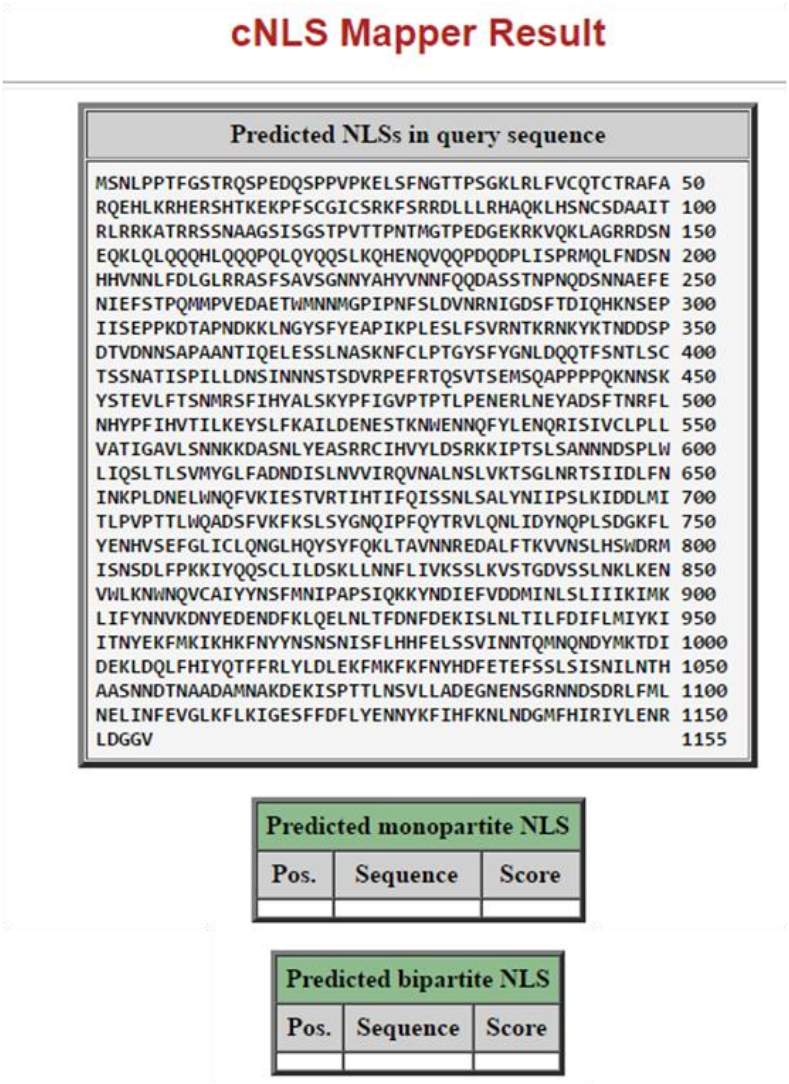


Figure A66: Prediction of importin α-dependent nuclear localisation signals in Mxr1p by cNLS Mapper (Kosugi et al, 2009). No putative monopartite or bipartite cNLS's could be located within the full amino acid sequence for Mxr1p.